

**DEPARTAMENTO DE BIOLOGÍA MOLECULAR
FACULTAD DE CIENCIAS
UNIVERSIDAD AUTÓNOMA DE MADRID**

ESTRUCTURA DE FIBRAS DE BACTERIÓFAGOS

Carmela García Doval

TESIS DOCTORAL

Madrid, 2013





**DEPARTAMENTO DE BIOLOGÍA MOLECULAR
FACULTAD DE CIENCIAS
UNIVERSIDAD AUTÓNOMA DE MADRID**

ESTRUCTURA DE FIBRAS DE BACTERIÓFAGOS

Carmela García Doval

Licenciada en Biología

Director: Mark J. van Raaij

**DEPARTAMENTO DE ESTRUCTURA DE MACROMOLÉCULAS
CENTRO NACIONAL DE BIOTECNOLOGÍA (CNB-CSIC)
MADRID, 2013**



El Dr. Mark J. van Raaij, científico titular del Departamento de Estructura de Macromoléculas del Centro Nacional de Biotecnología del Consejo Superior de Investigaciones Científicas (CNB-CSIC) certifica que:

Doña Carmela García Doval, Licenciada en Biología por la Universidad de Santiago de Compostela, ha realizado bajo su dirección el trabajo titulado "Estructura de Fibras de Bacteriófagos" en el Departamento de Estructura de Macromoléculas del Centro Nacional de Biotecnología del Consejo Superior de Investigaciones Científicas (CNB-CSIC).

En su opinión, este trabajo reúne las condiciones exigidas por la legislación vigente y tiene la originalidad, el rigor y la calidad científica necesarios y suficientes para ser presentado para optar al Grado de Doctor en Biología por la Universidad Autónoma de Madrid.

Para que así conste y con los efectos oportunos, firma el presente certificado.

Dr. Mark J. van Raaij
Científico Titular

Esta tesis doctoral se ha desarrollado en el Centro Nacional de Biotecnología bajo la dirección del Dr. Mark J. van Raaij, y ha sido financiada con una beca predoctoral del programa de Formación de Profesorado Universitario (FPU) con la referencia AP2008-01412, del Ministerio de Educación. Además, se disfrutó de una estancia en el laboratorio de la Dra. Pascale Boulanger en el Institut de Biochimie et Biophysique Moléculaire et Cellulaire (IBBMC-CNRS) gracias al Proyecto Conjunto CSIC-CNRS titulado "Estructura y función de las fibras laterales del bacteriófago T5" con la referencia 2011FR0016, financiado por el Consejo Superior de Investigaciones Científicas (CSIC).

Agradecimientos

Esta tesis ha sido posible gracias al Dr. Mark J. van Raaij, al que querría expresar mi más sincero agradecimiento por todo lo que me ha enseñado durante estos últimos años. Para mi ha sido un gran director de tesis, pero además de eso me ha transmitido su gran pasión por la ciencia, por la cristalografía, y se ha convertido en más que un jefe, en un amigo que espero mantener en el futuro. Mark siempre ha confiado en mi como nadie había hecho antes, y además me ofreció la posibilidad de venirme con él a Madrid, algo por lo que siempre le estaré agradecida.

También querría agradecer todo el apoyo recibido por mis compañeros de laboratorio a lo largo de estos años; a aquellos que me acompañaron en Santiago: Pablo, Sergio, José y Patricia; y a los que poco a poco se fueron incorporando en Madrid: Laura D, Meritxell, Abhi, Laura C, Marta y Thanh. Todos ellos me han hecho la vida más amena durante este tiempo. También a los laboratorios cercanos, aquellos que nos han prestado ayuda en los momentos de cambio: el de Javier Benavente en Santiago y los de Pepe, JoseMari, JR y Carmen en nuestro departamento ya en Madrid. En este punto, tengo que agradecer a Ana Cuervo todo lo que me ha enseñado sobre fagos y microscopía electrónica, y desearle toda la suerte del mundo en su aventura cristalográfica. También a Ana y JR por el esfuerzo de leer esta tesis y aconsejarme, y a Florencio Pazos por la imagen de la portada, y por haberme enseñado a usar pov-ray, algún día seré capaz de hacer imágenes como las suyas.

Por otra parte, no puedo dejar de agradecer su acogedor recibimiento a aquellos laboratorios en los que he hecho estancias cortas. Guardo un especial recuerdo de Purdue y de toda la gente que allí conocí, sin duda a quien más tengo que agradecer es al Prof. Michael Rossmann, que además de acogerme en su laboratorio y tratarme como una más de su gran familia, me permitió aprender más allá de los programas, y se convirtió en un modelo a seguir, transmitiéndome su ilusión por ver una proteína resuelta. También guardo buen recuerdo de Washington DC, del Prof. Venigalla Rao, que seguramente es la persona que más me ha enseñado sobre eficiencia, productividad y organización. En su grupo me sentía como en casa, y trabajar ese tiempo mano a mano con Zahid me permitió descubrir que clase de postdoc quiero ser algún día, además de enseñarme todo lo que sé sobre crecimiento de fagos y selección de mutantes. Por último, mi paso por el laboratorio de la Dra. Pascale Boulanger en París, del que también guardo buenos recuerdos, y fue la estancia con la que se inició el

trabajo con la fibra de T5, que forma parte de esta tesis. De la misma forma, también querría agradecer los buenos ratos a todos aquellos que he conocido en cursos y congresos, con los que he compartido experiencias que me han hecho crecer como investigadora, en especial a mis compañeros del HERCULES, con los que aprendí que en un sincrotrón se hacen cientos de cosas más allá de la cristalografía de proteínas.

Llegar hasta aquí no ha sido fácil, y no habría sido posible si a lo largo de mi vida no hubiera tenido el apoyo de gente que creyó en mí y me apoyó en las decisiones que el resto consideraban una locura, ya fuese estudiar Biología, una carrera que todos tildaban como "sin futuro" o empezar un doctorado. En este aspecto, recuerdo con especial cariño a algunos profesores del instituto, que pese a que el ambiente no era el más favorable, me animaron a que siguiera adelante con mis planes de futuro tan aparentemente suicidas. También de la facultad guardo grandes recuerdos, especialmente de un profesor de genética, Gonzalo, que viendo nuestro interés por los genes nos dijo que estudiásemos proteínas, espero poder hacerle llegar que al menos una alumna sí siguió su consejo. Y por supuesto a mis compañeros de facultad, en concreto querría agradecer a Laura y a David, que siempre me animaban cuando perdía fuerzas.

A lo largo de esta tesis mi vida ha cambiado mucho, y quizá el mayor cambio fue el traslado a Madrid, dejando mi querida Galicia (que nunca se sabe lo querida que es hasta que se deja) y mudándome a una gran ciudad rodeada de desconocidos. Pese a la dificultad del momento, siempre tuve a mis amigos que, aunque lejanos geográficamente, siempre he considerado cercanos animándome, y que en cada reunión "hamburguesil" me preguntaban por mi tesis, siempre interesándose por lo que hacía aunque el tema les pareciese extraño, y aunque tuviese que acabar haciendo dibujos de electrones dando vueltas en una servilleta.

También quiero agradecer a mi familia el apoyo que me han brindado en todo momento, desde el momento en el que decidí que "de mayor" quería estudiar biología, hasta el día que les comuniqué que me pretendía ir al extranjero de postdoc para continuar con mi carrera científica.

Y por último, pero no por ello menos importante, quiero agradecer a Nacho que me invitase a aquel café en aquella terraza, que hablase de Grenoble, y todo lo que ha venido después durante estos tres años.

Resumen

Los bacteriófagos son virus que infectan bacterias. Para ello, después de reconocer a la célula hospedadora y transferir el material genético a su interior y redirigen toda la maquinaria celular para producir más partículas virales. El primer paso fundamental para que este proceso se pueda llevar a cabo es la correcta interacción entre el fago y la membrana de la célula, por lo que los virus han desarrollado mecanismos para realizar eficientemente este proceso.

La mayor parte de los bacteriófagos descritos pertenecen al orden *Caudovirales*, que se caracterizan por tener una cápside icosaédrica o prolada y una cola. En muchos casos, en la base de la cola hay fibras o espinas, que actúan como receptores de la célula hospedadora. Las fibras son proteínas triméricas que en su extremo distal presentan el dominio de unión al hospedador. Son proteínas muy estables, pero su forma de L también les confiere también cierta flexibilidad, necesaria para su unión al receptor, generalmente reversible.

Los *Caudovirales* agrupan tres familias diferenciadas en función de la morfología de su cola: *Myoviridae*, *Podoviridae* y *Siphoviridae*. La familia *Myoviridae* incluye fagos con una cola larga y contráctil; siendo T4 el myovirus mejor estudiado, que tiene seis fibras largas y seis fibras cortas. La familia *Podoviridae* abarca los fagos con una cola corta y no contráctil. A esta familia pertenece el fago T7, que tiene seis fibras formadas por trímeros de la proteína gp17. Estas fibras han sido caracterizadas únicamente por microscopía electrónica y análisis bioquímicos. La familia *Siphoviridae*, está formada por fagos con una cola larga y flexible. El siphovirus T5 tiene tres fibras largas con forma de L formadas por trímeros de la proteína pb1, que han sido poco estudiadas hasta el momento.

En esta tesis doctoral se planteó el estudio de la estructura del extremo carboxilo terminal de las fibras de los fagos T7 y T5, formadas por las proteínas gp17 y pb1 respectivamente. La determinación de la estructura atómica de los extremos de las fibras de los fagos abre nuevas posibilidades para el desarrollo de aplicaciones biotecnológicas, como la modificación de los aminoácidos que interaccionan con el receptor y, por lo tanto, redireccionando los fagos hacia otras bacterias de interés. Debido a la aparición, cada vez más frecuente, de bacterias con resistencia a múltiples antibióticos, los fagos se presentan como una posible alternativa contra cepas bacterianas que causan patologías importantes en humanos.

Abstract

Bacteriophages are viruses that infect bacteria. For this infection they recognise the host cell, transfer the genetic material inside the cell and redirect the cell machinery in order to produce many viral particles. The first step for a successful infection is the correct interaction between the phage and the cell membrane. For this reason, viruses have developed mechanisms to bind efficiently to a specific receptor.

Most bacteriophages belong to the order *Caudovirales*. These phages have icosahedral or prolate capsids with a tail. In most cases, fibres or spikes are attached to the base of the tail, and function as receptors for the host cell. Fibres are trimeric proteins with a carboxy-terminal host-binding domain. Because of their folding, they are very stable proteins, resistant to denaturing agents and high temperature, but their L shape confers them flexibility, facilitating receptor binding. This receptor binding is generally reversible and is followed by the irreversible interaction with a secondary receptor. After the receptor binding, the tail penetrates through the cell membrane and the phage DNA is ejected into the cell cytoplasm.

The *Caudovirales* are divided into three families according to their tail morphology. The *Myoviridae* family consists of phages with a long contractile tail. T4, perhaps the best-known member of the *Myoviridae* family, has six long tail fibres and six short tail fibres. The *Podoviridae* family consists of phages with a short non-contractile tail. Bacteriophage T7, which has six fibres made by gp17 trimers, belongs to this family. These fibres have been previously analysed by electron microscopy and biochemical studies, but the atomic structure was still unknown. The last family, the *Siphoviridae* family, consists of phages with a long flexible tail. The siphovirus T5 has three L-shaped long fibres made by trimers of protein pb1 and they have been poorly studied.

In this PhD thesis, a study of the carboxy-terminal parts of T7 and T5 fibres, made by the proteins gp17 and pb1 respectively, has been done. The knowledge of the atomic structure of the distal part of the fibres opens new approaches to future biotechnological applications: amino acids interacting with the receptor can be modified to redirect the phage to other bacteria of interest. At a time when the antibiotics are losing their effectiveness to resistant bacteria, phages represent a possible alternative.

Índice

Resumen	1
Abstract	3
Índice	5
Abreviaturas	9
1 Introducción	11
1.1 Origen y diversidad estructural de los virus	13
1.2 Los bacteriófagos	15
1.3 Ciclo de replicación y patogenicidad	17
1.4 Características de la pared celular bacteriana.....	18
1.5 El bacteriófago T4 de la Familia Myoviridae.....	21
1.6 El bacteriófago T5 de la Familia Siphoviridae.....	24
1.7 El bacteriófago T7 de la Familia Podoviridae	26
1.8 Aplicaciones de los bacteriófagos	28
1.8.1 Terapia fágica	28
1.8.2 <i>Phage display</i>	30
1.8.3 Tipado bacteriano	30
1.8.4 <i>Reporter phages</i>	31
1.9 Conclusión	31
2 Objetivos.....	33
3 Materiales y métodos.....	37
3.1 Métodos biológicos	39
3.1.1 Cepas bacterianas	39
3.1.2 Plásmidos.....	40
3.1.3 Medios de cultivo	41
3.2 Manipulación de células bacterianas	42
3.2.1 Obtención de células competentes.....	42
3.2.2 Transformación química bacteriana	42
3.3 Manipulación de ácidos nucleicos.....	42
3.3.1 Amplificación del ácido nucleico	42
3.3.2 Corte con enzimas de restricción.....	43
3.3.3 Aislamiento de fragmentos de DNA y purificación	43
3.3.4 Purificación del DNA plasmídico.....	43
3.3.5 Ligación y obtención de vectores	43
3.4 Manipulación de proteínas.....	44

Estructura de fibras de bacteriófagos

3.4.1	Análisis por SDS-PAGE.....	44
3.4.2	Análisis por Tricina-SDS-PAGE.....	44
3.5	Purificación mediante cromatografía	44
3.5.1	Cromatografía de afinidad	44
3.5.2	Cromatografía de intercambio iónico	45
3.5.3	Cromatografía de exclusión molecular	45
3.5.4	Cromatografía de interacción hidrofóbica	45
3.5.5	Cromatografía en columna de hidroxipatito	46
3.6	Técnicas cristalográficas	46
3.6.1	Naturaleza de los cristales de proteína	46
3.6.2	Proceso de cristalización	47
3.6.3	Recogida de datos	49
3.7	Resolución de la estructura	50
3.7.1	El problema de la fase	50
3.7.2	Dispersión anómala simple/múltiple	51
3.7.3	Reemplazamiento molecular	52
3.7.4	Construcción del modelo	52
3.7.5	Validación.....	52
3.8	La proteína gp17	53
3.8.1	Clonación del gen de gp17	53
3.8.2	Clonación de los fragmentos carboxilo terminales de gp17.....	53
3.8.3	Expresión de gp17 sin cola de histidinas	54
3.8.4	Purificación de gp17 sin cola de histidinas	54
3.8.5	Expresión y purificación de gp17 con cola de histidinas	55
3.8.6	Expresión de los fragmentos carboxilo terminales de gp17 sin cola de histidinas.....	55
3.8.7	Purificación de los fragmentos carboxilo terminales gp17(371-553) y gp17(431-553) sin cola de histidinas	56
3.8.8	Expresión y purificación del fragmento gp17(371-553) con cola de histidinas...	56
3.8.9	Tripsinización de gp17 y del fragmento gp17(371-553).....	56
3.8.10	Microscopía electrónica de gp17	57
3.8.11	Cristalización del fragmento gp17(371-553).....	57
3.9	Ensayos de inhibición del fago T7	59
3.10	Purificación de colas del bacteriófago T5	61
3.10.1	Crecimiento y purificación de fagos (mutante <i>T5amD20</i>)	61
3.10.2	Purificación de fagos mediante gradiente de cloruro de cesio	62

3.10.3	Crecimiento y purificación de colas	62
3.10.4	Microscopía electrónica y procesamiento digital de las colas de T5	62
3.11	La proteína pb1	63
3.11.1	Predicción de estructura secundaria de la fibra de T5	63
3.11.2	Clonación de pb1 y de los fragmentos carboxilo terminales.....	64
3.11.3	Purificación de los fragmentos pb1(794-1264), pb1(1106-1264) y pb1(970-1264).....	65
3.11.4	Expresión y purificación del fragmento pb1(970-1264) con selenio-metionina.....	65
3.11.5	Cristalización de los fragmentos de pb1	66
3.12	Alineamiento de secuencias	67
3.12.1	La fibra de T7	67
3.12.2	La fibra de T5	67
4	Resultados y discusión	69
4.1	La proteína gp17	71
4.1.1	Clonaciones	71
4.1.2	Expresión y purificación de los fragmentos de gp17	72
4.1.3	Tripsinización de gp17	75
4.1.4	Microscopía electrónica de las fibras de T7	76
4.1.5	Cristalización del fragmento gp17(371-553).....	77
4.1.6	Descripción de la estructura	82
4.1.7	El dominio pirámide	83
4.1.8	El dominio punta	85
4.1.9	Estabilidad y plegamiento de gp17(371-553).....	86
4.1.10	Unión a receptor	87
4.2	Ensayos de inhibición de la infección por T7	89
4.3	Comparación de gp17 con otras proteínas	94
4.4	La cola del bacteriófago T5	100
4.4.1	Purificación de colas de T5	100
4.4.2	Microscopía electrónica de las fibras de la cola de T5	101
4.5	La proteína pb1 del bacteriófago T5	102
4.5.1	Predicción de estructura secundaria	102
4.5.2	Clonaciones de fragmentos de pb1	105
4.5.3	Expresión de los fragmentos de pb1	105
4.5.4	Purificaciones de fragmentos de pb1	106
4.5.5	Cristalización del fragmento pb1(794-1264).....	110

4.5.6	Cristalización del fragmento pb1(970-1264).....	111
4.5.7	Descripción de la estructura	113
4.5.8	Estabilidad y plegamiento	117
4.5.9	Unión al receptor	119
4.5.10	Superposición de la estructura	120
4.6	Comparación de pb1 con otras estructuras	121
4.6.1	Alineamiento de secuencias.....	123
4.7	Perspectivas	126
5	Conclusiones	127
6	Bibliografía	131
	Anexo 1: Tabla de oligonucleótidos	147
	Anexo 2: Índice de figuras	149
	Anexo 3: Índice de tablas	151
	Anexo 4: Artículos publicados durante el desarrollo de la tesis	153

Abreviaturas

BTV: virus de la lengua azul
DO₆₀₀: densidad óptica a 600 nm
EPS: exopolisacáridos
FhuA: Componente de captación de hierro A
gp (referido a proteínas): producto génico
IPTG: Isopropil β -D-tiogalactósido
HK97: virus Hong Kong 97
Hoc: proteína externa de la cápside altamente antigénica
LB: medio de cultivo Luria-Bertani
ltf: fibras largas de la cola (referido a T4)
lrf: fibras de la cola con forma de L (referido a T5)
MME: mono metil éter
MOI: multiplicidad de infección
Ni-NTA: níquel-ácido nitriloacético
Omp: proteína de la membrana externa
PDB: base de datos de proteínas (*Protein Data Bank*)
PEG: polietilenglicol
pb: pares de bases
pb: banda de proteína (referido a proteínas)
RBP: proteína de unión a receptor
Soc: proteína pequeña externa de la cápside
STIV: virus icosaédrico con torreta de Sulfolobus
stf: fibras cortas de la cola (referido a T4)
str: fibra recta de la cola (referido a T5)

1 Introducción

1.1 Origen y diversidad estructural de los virus

Se estima que en la biosfera hay aproximadamente 10^{31} - 10^{32} partículas virales (Bergh *et al.*, 1989), que son las responsables de la mayor presión selectiva sobre el resto de organismos (Bamford *et al.*, 2005). Esto implica que existe una gran diversidad viral, desde los circovirus porcinos con sólo dos genes (Allan *et al.*, 2012) hasta los mimivirus, con más de 900 genes (Claverie *et al.*, 2009). El origen de esta gran diversidad viral no ha sido dilucidado, pero existen principalmente tres hipótesis sobre su origen:

- Los virus se originaron antes de la vida celular.
- Los virus se originaron a partir de las células por un proceso de reducción.
- Los virus se liberaron de las células, pero continuaron utilizando sus mecanismos de replicación.

Esta última hipótesis no explica cómo se podrían formar estructuras tan avanzadas como los complejos para el empaquetamiento de DNA viral, que no tienen homólogos en las estructuras celulares, pero de todas formas tampoco existe un consenso en la comunidad científica sobre cual de las otras dos hipótesis es la correcta (Abrescia *et al.*, 2012).

Desde el descubrimiento de los virus, se empezaron a postular muchas hipótesis sobre su posible estructura. En 1935 se cristalizó el primer virus, el virus del mosaico del tabaco (TMV) (Stanley, 1935). En 1978 se obtuvo la primera estructura viral a alta resolución, la del virus del enanismo ramificado del tomate (TBSV) (Harrison *et al.*, 1978), y poco después la del virus del mosaico sureño de la judía (SBMV) (Abad-Zapatero *et al.*, 1980). La estructura del SBMV mostró por primera vez que las subunidades que forman la cápside son similares entre virus diferentes, lo que sugería que muchos virus están relacionados entre sí. La primera prueba de esta relación, incluso entre virus que infectaban a hospedadores diferentes, se obtuvo con la primera estructura determinada de un virus humano, el rinovirus (Rossmann *et al.*, 1985), cuya estructura es homóloga a la de algunos virus de plantas. Posteriormente también se descubrió que los bacteriófagos están relacionados con los herpesvirus dada la similitud en la estructura de algunas de sus proteínas y por su forma de ensamblarse (Steven *et al.*, 2005), y por estas características se ha definido que forman parte de un mismo linaje viral. El hecho de que distintos virus de un mismo linaje (como los bacteriófagos y los herpesvirus) infecten hospedadores de distintos

dominios, puede ser debido a que los antecesores comunes para cada uno de estos linajes existiesen antes de la separación de los diferentes dominios de la vida. Además, hay que tener en cuenta que las similitudes que definen un linaje no se limitan a las proteínas de la cubierta, ya que también existen similitudes en la estructura de su genoma o el mecanismo de encapsidación (Bamford *et al.*, 2005). Alternativamente, también existe la posibilidad de que los virus se hayan adaptado a nuevos hospedadores, similares o de dominios diferentes.

Considerando las estructuras atómicas de las proteínas de la cápside descritas hasta el momento, se pueden diferenciar cuatro linajes de virus (Bamford *et al.*, 2005) (Figura 1.1):

- Virus icosaédricos con DNA bicatenario del tipo PRD1 o tipo Adenovirus: en este grupo se incluye por ejemplo el virus STIV (virus icosaédrico con torreta de *Sulfolobus*, que infecta a arqueas).
- Virus con DNA bicatenario con cola que infectan bacterias y arqueas, y herpesvirus: virus tipo HK97 (Hong Kong 97).
- Virus con RNA bicatenario: el virus más representativo es el virus de la lengua azul (BTV).
- Virus tipo Picornavirus.

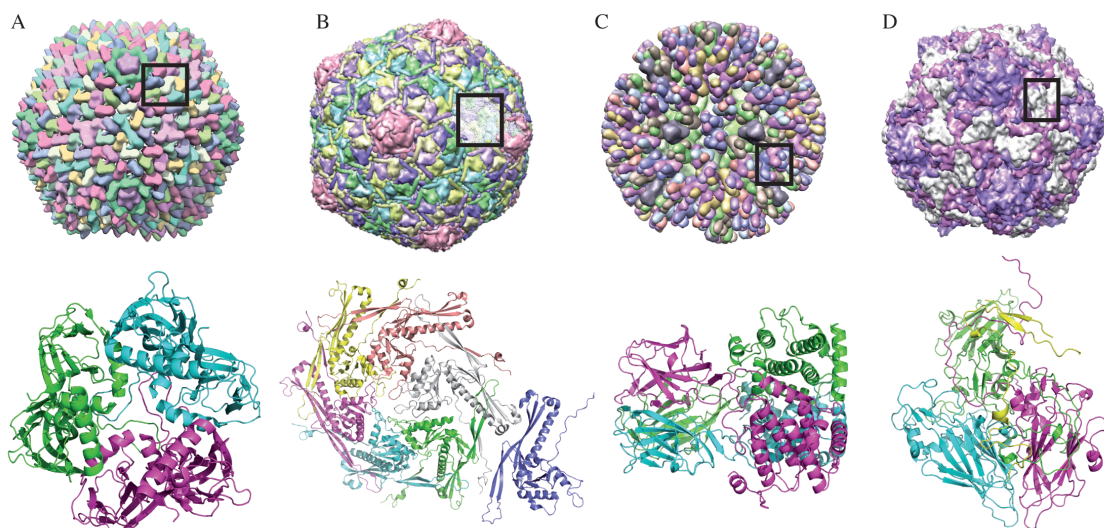


Figura 1.1: Virus pertenecientes a los cuatro linajes descritos según la similitud de sus proteínas de la cápside. (A) PRD1, (B) HK97, (C) BTV, (D) ECV-1 (Echovirus-1). En la fila superior se representa la cápside y en la fila inferior la estructura cristalográfica de la proteína (o complejo de proteínas) mayoritaria de la cápside, siendo evidente que pese a la similitud externa en la cápside, la organización de sus proteínas es diferente en cada caso.

1.2 Los bacteriófagos

Los bacteriófagos son virus que infectan bacterias. Al igual que el resto de los virus, reconocen específicamente a sus bacterias hospedadoras, les transfieren su material genético y redirigen la maquinaria de la célula para producir más partículas virales, habiendo desarrollado diferentes mecanismos para llevar a cabo esta infección. Los bacteriófagos fueron descubiertos de forma paralela a principios del siglo XX por Frederick William Twort (Twort, 1915) y Felix Hubert d'Herelle (d'Herelle, 1917), aunque su actividad lítica ya había sido observada previamente por Ernst Hankin en 1896 (Hankin, 1896). Twort llegó a la conclusión de que la formación de calvas en las placas de cultivo de bacterias se debía a agentes infecciosos no filtrables, que mataban a las bacterias y que se multiplicaban durante el proceso de infección. D'Herelle, que trabajaba con disentería, observó que había un agente al que llamó "microbio invisible", que era capaz de matar a las bacterias y que se encontraba en el intestino de las personas recuperadas, pero no en el de personas enfermas o sanas. Este agente podía filtrarse y seguía siendo activo, y se podía transmitir, por lo que tenía que ser alguna clase de germen vivo, que no crecería en un medio artificial, siendo un parásito obligatorio, al que él mismo denominó bacteriófago (Duckworth, 1976).

Los bacteriófagos infectan una gran variedad de bacterias y se localizan en cualquier medio: suelos, agua dulce o salada, animales y plantas. La mayor parte de los fagos identificados hasta el momento se pueden cultivar fácilmente en bacterias de interés industrial o médico (Ackermann y Kropinski, 2007) (Brussow y Hendrix, 2002), lo que ha simplificado su caracterización. Por otro lado, esto sugiere que la mayor parte de fagos que todavía no han sido identificados, probablemente no se puedan cultivar fácilmente. La mayor parte de los bacteriófagos identificados pertenecen al orden *Caudovirales*, siendo fagos con cola y DNA bicatenario lineal. Considerando su forma, además de caudovirales hay fagos icosaédricos, helicoidales y pleomórficos. El ácido nucleico puede ser DNA o RNA, monocatenario o bicatenario, y puede ser superenrollado o segmentado. También hay fagos que tienen una envuelta lipídica. Por lo tanto, aunque los caudovirales sean los fagos más abundantes en la naturaleza, existe una gran diversidad morfológica (Van Regenmortel *et al.*, 2000).

Estructura de fibras de bacteriófagos

Los bacteriófagos pertenecientes al orden *Caudovirales* pertenecen al linaje HK97. Poseen una cápside icosaédrica o prolada cuyo radio varía entre 10 y 100 nm. En el interior de la cápside se encuentra el DNA, con un tamaño de 17 a 725 kb (Ackermann, 1998). La cola es un tubo hueco helicoidal, generalmente con simetría de orden seis, que contiene estructuras adicionales como la placa basal, y en algunos casos proteínas alargadas, que pueden ser espinas o fibras. Aunque no todos los fagos tienen fibras, todas las familias tienen representantes con fibras. Estas fibras intervienen en el reconocimiento de la célula hospedadora, y la mayor parte de ellas comparten una organización estructural similar: un extremo amino terminal por el que se unen al virus, un dominio "asta" central, de longitud variable, y una región globular carboxilo terminal, en la que normalmente se encuentra el dominio de unión al receptor.

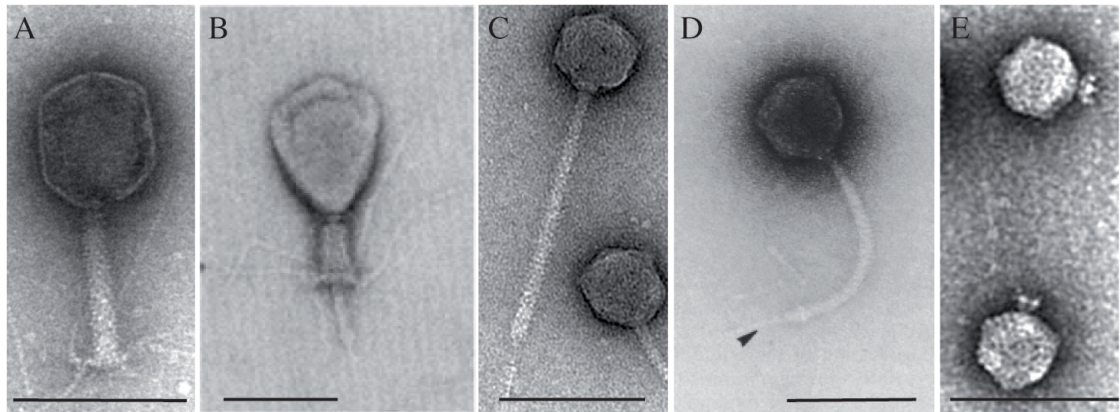


Figura 1.2: Imágenes de microscopía electrónica de representantes de cada una de las tres familias de Caudovirales. (A) Myovirus T4 con la cola extendida, (B) myovirus T4 con la cola contraída, (C) siphovirus HK97, (D) siphovirus T5, observándose que la cola es flexible, (E) podovirus P22. Las barras indican 100 nm. Imágenes adaptadas (Heller y Schwarz, 1985; Mesyanzhinov, 2004; Krupovic *et al.*, 2011)

Los *Caudovirales* se dividen en tres familias en función de la morfología de su cola (Ackermann y Kropinski, 2007) (Figura 1.2):

- *Myoviridae*: son fagos con una cola larga, no flexible y contráctil. A esta familia pertenecen fagos como T4 o μ . Representan una cuarta parte de los caudovirales descritos.

- *Siphoviridae*: son fagos con una cola larga, flexible y no contráctil. A esta familia pertenecen λ o T5. Representan aproximadamente la mitad de los caudovirales conocidos.

- *Podoviridae*: son fagos con una cola corta y no contráctil. Algunos de los representantes más estudiados son $\phi 29$ o T7. Suponen aproximadamente el cuarto restante de los caudovirales identificados.

1.3 Ciclo de replicación y patogenicidad

Según el tipo de infección y el ciclo de replicación, los caudovirales también se pueden dividir en dos grupos, pudiendo ser lisogénicos (o atemperados) y líticos. Los fagos lisogénicos integran su DNA en el genoma de la célula hospedadora o en forma de plásmido, y así, cada vez que la bacteria se divida, también se replicará el material genético del virus (Figura 1.3). Si las condiciones ambientales dejan de ser favorables para la bacteria, el virus inicia su ciclo lítico, liberando la progenie al medio (Orlova, 2012). Cuando los fagos permanecen integrados en la bacteria se denominan profagos, y actúan disminuyendo la competencia de la bacteria, ya que ésta tiene que replicar más material genético y además existe el peligro de que el fago pase a ciclo lítico. Sin embargo, la adquisición de profagos es una manera de generar variabilidad en las bacterias, ya que la adquisición y pérdida de profagos es un proceso mucho más rápido que el de especiación (Brussow *et al.*, 2004). Los fagos también son utilizados por las bacterias como “medio de transporte” para la transferencia genética, como en el caso de las islas de patogenicidad (Brussow *et al.*, 2004; Novick *et al.*, 2010). De este modo desempeñan un papel imprescindible en la emergencia de nuevas cepas patógenas, como el caso de *Escherichia coli* (*E. coli*) O157 (Ohnishi *et al.*, 2001), o en cepas portadoras de toxinas como difteria, botulismo o escarlatina, sin ninguna función en el ciclo del fago, pero proporcionando una ventaja a la bacteria (Brussow *et al.*, 2004). Con el paso del tiempo, la presión selectiva hace que muchos de estos profagos pierdan la capacidad de pasar al ciclo lítico y en muchos casos desaparecen por deleciones de las bacterias, quedando sólo aquellos fragmentos que son beneficiosos para la bacteria hospedadora (Calendar y Abedon, 2005).

Los fagos con ciclo lítico, al eyectar su DNA al interior de la célula inducen un cambio en la maquinaria celular, orientándola a la formación de nuevos virus, normalmente entre 50 y 200, para a continuación lisis la célula y liberar la progenie (Figura 1.3). La generación de una progenie que pueda infectar nuevas células va a depender en un primer paso de la capacidad del fago de unirse a la membrana.

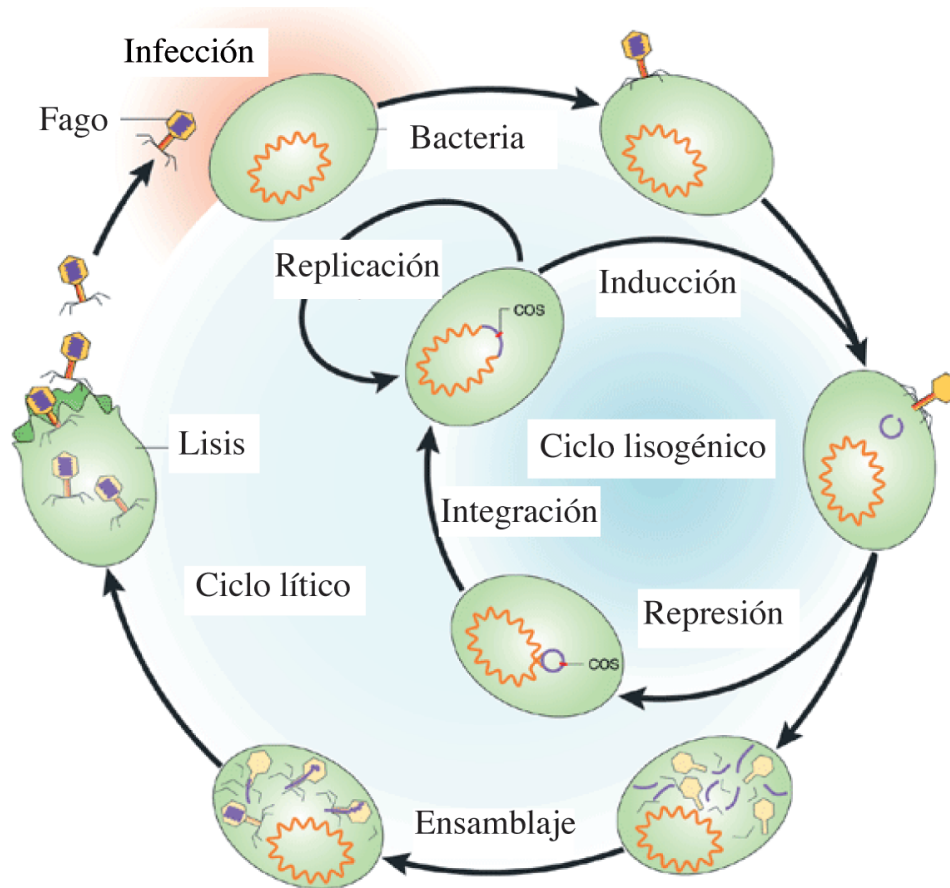


Figura 1.3: Ciclo de replicación de un fago. Tras el proceso de infección, el fago puede reprimirse, integrándose en el genoma (ciclo lisogénico) y replicándose con la célula hasta que se produzca la inducción y pase a ciclo lítico, o entrar directamente a esa fase, en la que generan las proteínas fágicas, se ensamblan y se lisa la célula liberando la progenie. Imagen adaptada (Campbell, 2003).

1.4 Características de la pared celular bacteriana

La mayor parte de las bacterias poseen una pared celular que da forma y rigidez a la célula. Las bacterias se pueden clasificar aplicando la tinción de Gram en función de la estructura de su pared, pudiendo ser Gram positivas o Gram negativas. La tinción de Gram consiste en una primera tinción con cristal violeta, que se une a la parte externa de la célula, una fijación, una decoloración que elimina los lípidos de la membrana externa y una segunda tinción con safranina. Así, las bacterias Gram positivas quedarán teñidas de violeta, mientras que las Gram negativas perderán el violeta con la decoloración al eliminarse la membrana externa y se teñirán de rosa con la safranina.

La membrana citoplasmática o membrana interna está formada por una bicapa lipídica con proteínas transportadoras incluidas (Figura 1.4). En contacto con la membrana citoplasmática se encuentra el peptidoglicano o mureína, compuesto por repeticiones de polímeros de N-acetilglucosamina unida a ácido N-acetilmurámico por enlaces β -1,4 glucosídicos y con cadenas laterales formadas por tetrapéptidos, que están unidos directamente por enlaces interpeptídicos en la mayoría de las Gram negativas, o mediados por puentes de entre uno a siete aminoácidos en la mayoría de las Gram positivas. Mientras que en las bacterias Gram positivas la pared mide entre 15 y 80 nm y está formada por varias capas de peptidoglicano asociado a ácidos teicoicos; en las Gram negativas mide aproximadamente 10 nm y está compuesta por una única capa de peptidoglicano, rodeada por una membrana externa (Figura 1.4).

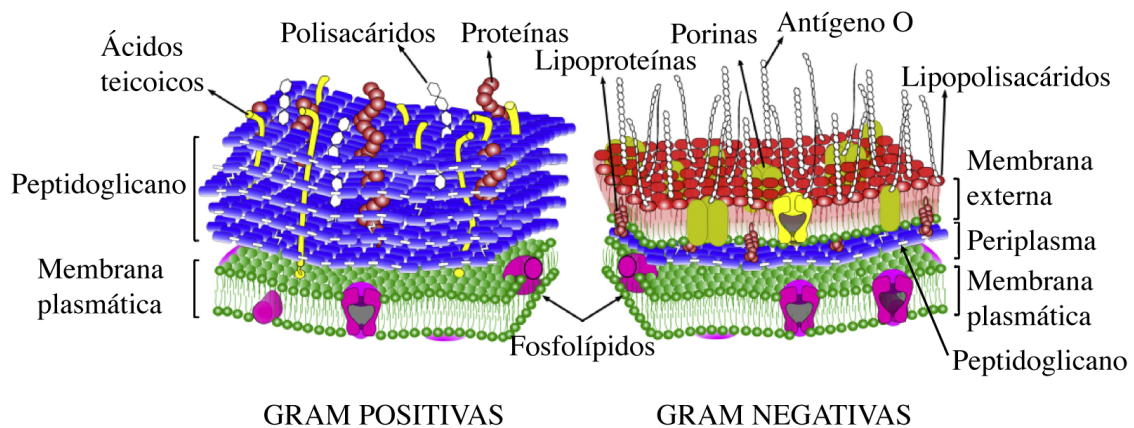


Figura 1.4: Esquema de la pared celular de bacterias Gram positivas y Gram negativas. En las Gram positivas (izquierda), la membrana plasmática se encuentra rodeada de varias capas de peptidoglicano que en su interior contiene ácidos teicoicos, polisacáridos y proteínas. En las Gram negativas (derecha), la capa de peptidoglicano es muy fina y está rodeada por la membrana externa, en la que quedan expuestos los lipopolisacáridos. Imagen adaptada (Tripathi *et al.*, 2012).

La membrana externa de las bacterias Gram negativas es una segunda bicapa lipídica formada por fosfolípidos, polisacáridos y proteínas (Figura 1.4). Entre las proteínas transmembrana, las más abundantes son las porinas, homotrómeros que forman un poro de aproximadamente 1 nm de diámetro, que permite la difusión de moléculas pequeñas. Las porinas también actúan como receptores para algunos fagos. La monocapa exterior de la membrana externa de la mayor parte de las

bacterias Gram negativas está formada principalmente por una molécula, denominada lípido A, que está formado por un esqueleto de 2,3-diacilglucosamina, lo que podría conferirle una mayor resistencia frente a fosfolipasas (Figura 1.5). El lípido A sirve de anclaje a la membrana para el resto de la cadena de lipopolisacáridos o endotoxinas (reciben este nombre por ser tóxicos para muchos animales). En los lipopolisacáridos, la posición 6' del lípido A está glucosilada con un oligosacárido, formando el núcleo (Figura 1.5). La mayor parte de los lipopolisacáridos de bacterias patógenas presentan además otro polisacárido, unido a la glucosa distal del núcleo, que se denomina antígeno O y consiste en una serie de repeticiones del mismo oligosacárido (Seydel *et al.*, 1989; Raetz, 1990) (Figura 1.5).

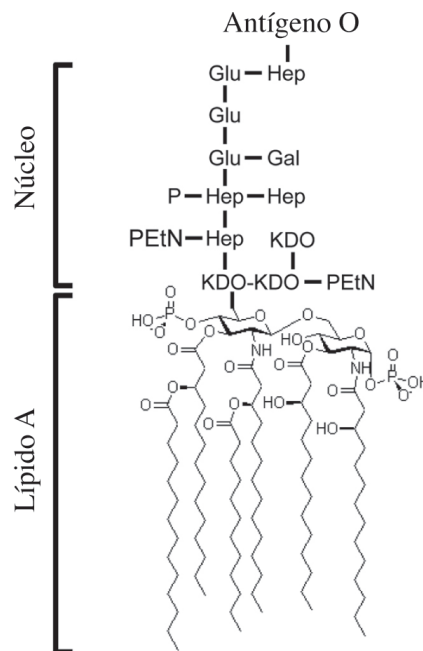


Figura 1.5: Esquema de la estructura de los lipopolisacáridos. Se pueden diferenciar tres regiones: el lípido A, el núcleo y el antígeno O, que no siempre está presente y muestra mayor variabilidad. Imagen adaptada (Ruiz *et al.*, 2009).

Las bacterias que tienen antígeno O se denominan lisas, y las que carecen de él, rugosas. Además de la estructura de los lipopolisacáridos, estos adjetivos se refieren también a la morfología de las colonias, ya que aquellas que no presentan antígeno O tienen los bordes rugosos. Aunque este antígeno confiere una ventaja evolutiva, para que una cepa pueda crecer en condiciones de laboratorio sólo es necesario que presente el lípido A y el núcleo, es decir, que sea rugosa, como las cepas más utilizadas en los laboratorios, *E. coli* K-12 y B. Los polisacáridos O son producidos

por las bacterias Gram negativas *Enterobacteriaceae*, *Pseudomonadaceae*, *Pasteurellaceae* y *Vibrionaceae*. La diversidad de antígenos O incluye más de sesenta monosacáridos y treinta componentes que no son carbohidratos (Raetz y Whitfield, 2002). La estructura del polisacárido O define la especificidad serológica del antígeno O en una bacteria, pero el número de variantes de antígeno O en una especie es variable. Por ejemplo, en *E. coli* hay aproximadamente 170 serotipos y en *Salmonella* se han descrito más del mil. Esta diversidad ayuda a la bacteria a evadir el sistema inmunológico (Raetz, 1990). La resistencia de las distintas cepas bacterianas a una batería de bacteriófagos se ha utilizado para determinar la estructura de los lipopolisacáridos (Boman y Monner, 1975). En la actualidad se utiliza una técnica basada en el mismo principio, pero en sentido inverso, para identificar las bacterias de una muestra mediante tipado bacteriano.

1.5 El bacteriófago T4 de la Familia Myoviridae

Los fagos pertenecientes a la familia *Myoviridae* tienen una cola larga y contráctil. El mecanismo contráctil de la cola permite hacer agujeros en la membrana externa de la célula y atravesar el periplasma y la membrana interna, para llevar a cabo la eyección del DNA. Para ello, el tubo externo de la cola se contrae, permitiendo que el tubo interno perfora la pared celular. En la punta del tubo interno está localizada una aguja molecular que ayuda a penetrar la membrana (Browning *et al.*, 2012).

El representante mejor estudiado de los *Myoviridae* es el bacteriófago T4. El fago T4 posee una cápside prolada (T=13l, Q=20), una cola contráctil y una placa basal con seis fibras cortas y seis fibras largas (Figura 1.6A). La cápside icosaédrica elongada (prolada) de T4 mide 120 nm de largo y 86 nm de ancho y está formada principalmente por dos proteínas: gp23 y gp24. Gp23 forma hexámeros y gp24 forma pentámeros en once de los doce vértices. El último vértice (vértice portal o conector) está formado por gp20, que es un dodecámero, y a través de este vértice se empaqueta el DNA y se produce su salida durante el proceso de infección (Rao y Black, 2010). Además de estas proteínas, la cápside posee otras dos proteínas no esenciales en su superficie: Hoc (proteína externa de la cápside altamente antigénica) y Soc (proteína pequeña externa de la cápside). Hoc se localiza en el centro de cada uno de los hexámeros, generando una protuberancia de unos 6 nm, y Soc está formando una malla entre los hexámeros de gp23 (Miernikiewicz *et al.*, 2012). La

presencia de Soc confiere al bacteriófago mayor estabilidad en ambientes hostiles como aquellos con alto pH, alta temperatura, choques osmóticos o agentes desnaturalizantes (Ishii *et al.*, 1978). Hoc, que presenta dominios similares a las inmunoglobulinas, podría participar en la interacción con la superficie bacteriana manteniendo al fago en los alrededores de la célula hasta que realice el contacto con sus receptores, incrementando así las posibilidades de infección (Fraser *et al.*, 2007; Sathaliyawala *et al.*, 2010). El bacteriófago T4 pertenece al linaje del virus HK97, y las estructuras de las principales proteínas de su cápside, gp23 y gp24, son similares a la principal proteína de la cápside de HK97 (Rao y Black, 2010). En el vértice portal, junto con el dodecámero de gp20 se encuentran las proteínas de la terminasa viral gp17 y gp16, constituyendo la maquinaria de empaquetamiento del DNA (Bhattacharyya y Rao, 1993). Gp17 es la subunidad grande y lleva a cabo el empaquetamiento del DNA, mientras que gp16 es la subunidad pequeña y está implicada en el reconocimiento del DNA, además de regular la actividad de gp17 modulando la actividad ATPasa, translocasa y nucleasa (Al-Zahrani *et al.*, 2009). Estas dos últimas proteínas, gp16 y gp17 no forman parte del virus maduro.

La cola de T4 está formada por una vaina, un tubo interno y una placa basal situada en el extremo distal de la cola (Figura 1.6B). Las fibras cortas y las fibras largas se encuentran unidas a la placa basal, y son responsables del reconocimiento y unión a la célula hospedadora. La vaina y el tubo interno están formados por 138 subunidades de gp18 y gp19 respectivamente, y se unen por su extremo proximal al complejo del cuello, formado por gp3, gp15, gp13 y gp14, que une la cola con la cápside. A este complejo también se une la fibritina, una proteína fibrilar que forma los seis "bigotes" (*whiskers*) del fago (Leiman y Shneider, 2012; Fokine *et al.*, 2013). La placa basal es hexamérica y contiene al menos 16 proteínas diferentes, incluyendo seis trímeros de gp12 que forman las fibras cortas (*stf*, *short tail fibres*) (Thomassen *et al.*, 2003) (Figura 1.6C). Las fibras cortas y largas requieren a la chaperona gp57 durante el ensamblaje de las fibras, pero no forma parte de la partícula viral (Coombs y Arisaka, 1994).

Las fibras largas (*ltf*, *long tail fibres*) son estructuras en forma de L con una longitud aproximada de 145 nm y un ancho variable de hasta 5 nm, que se pueden dividir en una parte proximal y otra distal. La parte proximal está formada por un homotrímero de gp34 que se une por su extremo amino terminal a gp9 de la placa basal y por el carboxilo terminal a gp35, que forma la "rodilla" de la fibra. La parte

distal está formada por un homotrímero de gp36 y otro de gp37, que es la parte más distal y la que interacciona con el receptor (Leiman *et al.*, 2010). Únicamente la parte más distal de gp37 ha sido descrita estructuralmente (Bartual *et al.*, 2010b) (Figura 1.6D). Para el correcto plegamiento de gp34 y gp37, de la misma forma que para el de gp12, es necesaria la chaperona gp57 y en el caso de gp37 también se requiere la chaperona gp38 tampoco va a formar parte de la partícula vírica (Bartual *et al.*, 2010a).

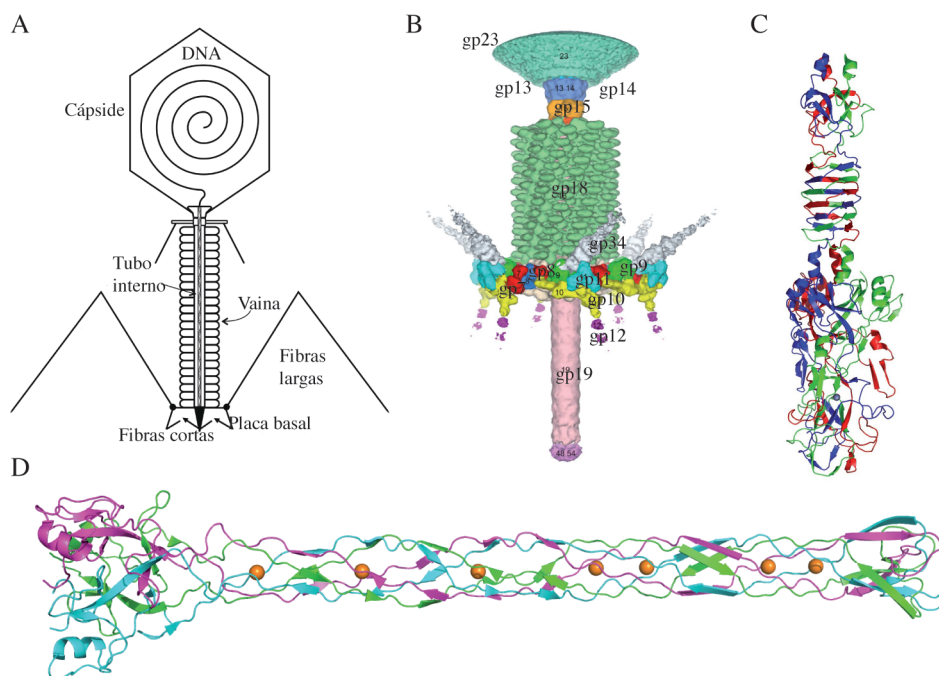


Figura 1.6: El myovirus T4. (A) representación esquemática de T4 con los principales elementos estructurales señalados, (B) reconstrucción de la cola de T4 contraída con las diferentes proteínas señaladas, (C) estructura cristalográfica de la parte carboxilo terminal de gp12, (D) estructura cristalográfica de la parte carboxilo terminal de gp37. Imagen adaptada (Leiman *et al.*, 2010).

El primer reconocimiento del receptor por parte de gp37 es reversible, uniéndose a residuos glucosil- α -1,3-glucosa de los lipopolisacáridos de la bacteria (Dawes, 1975) o, alternativamente, a OmpC (proteína de la membrana externa C) (Yu y Mizushima, 1982). Cuando al menos tres de las seis fibras largas se han unido a los receptores (Wood y Henninger, 1969), se produce una transducción de señal a la placa basal, que se abre cambiando la conformación hexagonal a una en forma de estrella. Este cambio de conformación causa la extensión de las fibras cortas y libera

su parte carboxilo terminal mediante la cual interaccionan con el núcleo interno de los polisacáridos de forma irreversible. A continuación, la estructura de la vaina se contrae y el tubo interno penetra en la envoltura celular. En el extremo distal del tubo se encuentra una aguja formada por un trímero de gp5. La actividad lisozima de esta proteína forma un agujero en la membrana facilitando la entrada del tubo y la liberación del DNA en el interior de la célula (Leiman *et al.*, 2010). En el extremo distal de gp5 existe otra proteína de menor tamaño, aunque aún no ha sido caracterizada (Kostyuchenko *et al.*, 2003).

1.6 El bacteriófago T5 de la Familia Siphoviridae

Los bacteriófagos de la familia *Siphoviridae* se caracterizan por tener una cola larga, flexible y no contráctil (Figura 1.7). A esta familia pertenecen el fago λ (muy utilizado en biología molecular), el fago SPP1 (ampliamente estudiado desde un punto de vista estructural) y el fago T5. El bacteriófago T5 tiene una cápside icosaédrica T= 13 de aproximadamente 80 nm de diámetro que contiene 121 kpb de DNA bicatenario, y una cola con una longitud de 250 nm. Frente a otros caudovirales presenta una característica peculiar: la simetría de orden 3 en el tubo de la cola, mientras que lo común es una simetría de orden 6 (Figura 1.7). Su genoma tiene interrupciones en las que se convierte en monocatenario en posiciones determinadas, así como una redundancia de 10 kpb en los extremos (Bujard y Hendrickson, 1973).

La cápside está formada mayoritariamente por pb8, que constituye 120 hexámeros y 11 pentámeros (Zweig y Cummings, 1973). En el vértice portal de la cápside se encuentra pb7 que forma un anillo dodecamérico y actúa como nexo entre la cápside y la cola. Además, en los hexámeros se une pb10, una proteína de decoración que al igual que en el caso de Hoc de T4 es dispensable (Effantin *et al.*, 2006).

En el interior de la cola se encuentra pb2. Esta proteína presenta una región de hélice superenrollada (*coiled coil*) que funciona como sensor para abrir la conexión entre la cápside y la cola permitiendo la eyección del DNA (Boulanger *et al.*, 2008).

La interacción de T5 con la célula comienza con una unión reversible de las fibras con forma de L (*lrf*), formadas por un trímero de pb1, al antígeno O de los lipopolisacáridos. Para que se produzca la unión, las células deben poseer los serotipos O8 u O9, que se caracterizan por tener repeticiones de polimanasas, siendo un trimanósido el más potente inhibidor de la unión de las fibras (Heller y Braun, 1982). Esta interacción no es imprescindible, pero la interacción de las fibras con los

lipopolisacáridos puede multiplicar por 15 la tasa de adsorción del fago, como se ha descrito con la cepa *E. coli* F (Lanni, 1958; Heller y Bryniok, 1984). Tras la unión de las fibras se produce la unión irreversible de pb5, localizada en el extremo distal de la cola formando la fibra recta (*stf*), a la proteína FhuA, que en *E. coli* está implicada en el transporte de hierro al interior de la célula (Heller, 1992). La presencia de FhuA es suficiente para que se inicie la eyección del DNA de T5 (Boulanger *et al.*, 1996). FhuA también actúa como receptor de otros siphovirus como T1 o $\phi 80$.

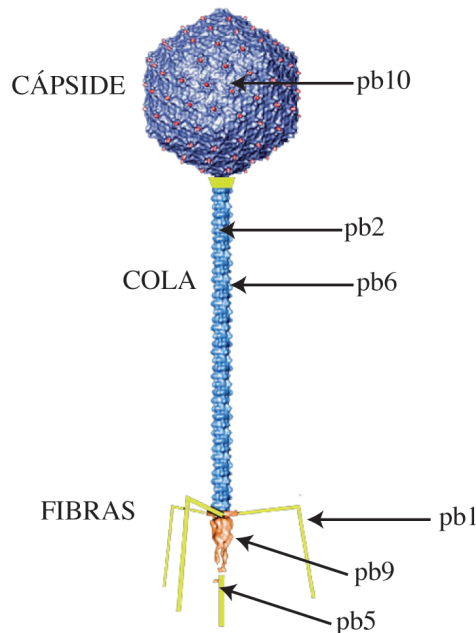


Figura 1.7: El siphovirus T5. Las principales proteínas estructurales están indicadas: pb10 en la cápside; pb2 (en el interior), pb6, pb9 y pb5 en la cola, y pb1 formando las fibras. Imagen adaptada (Effantin *et al.*, 2006)

El transporte del genoma de T5 al interior de la célula se realiza en dos pasos, dependientes de las características del genoma, por lo que no puede estar circularmente permutado como en el caso de T4. En un primer paso se eyecciona el 8% del genoma en el que se encuentran codificadas al menos nueve proteínas, entre las que se encuentran A1 y A2, necesarias para la eyección del resto del genoma en un segundo paso, aunque se ha descrito que algunas de estas proteínas no son esenciales en condiciones de laboratorio (McCorquodale *et al.*, 1977). Estas proteínas también inducen la degradación del DNA del hospedador e interrumpen la síntesis de proteínas celulares (Lanni, 1968). Tras una pausa, continúa la eyección del resto del

genoma, en el que se codifican el resto de proteínas necesarias para la replicación del genoma del fago y las proteínas estructurales (Snyder y Benzinger, 1981).

1.7 El bacteriófago T7 de la Familia Podoviridae

Los fagos pertenecientes a la familia *Podoviridae* presentan una cola corta no contráctil. A esta familia pertenecen algunos de los fagos mejor estudiados desde el punto de vista estructural y biofísico, como son P22, ϕ 29 y T7. El fago T7 se caracterizó en 1945 como miembro de los fagos tipo T que se replican de forma exclusivamente lítica en *E. coli* B (Demerec y Fano, 1945). T7 infecta cepas de *E. coli* rugosas (aquellas que no presentan la cadena del antígeno O en sus lipopolisacáridos).

La cápside icosaédrica T=7, de unos 60 nm de diámetro, está formada por gp10, el conector y el núcleo. Dentro de la cápside, el DNA de 40 kpb está ordenado formando seis capas coaxiales (Cerritelli *et al.*, 1997). El conector dodecamérico une la cápside, el núcleo y la cola, y está formado por gp8 (Serwer, 1976; Serwer *et al.*, 1997). El núcleo de T7 es una estructura cilíndrica, de 26 nm de largo y 21 nm de ancho, formada por cuatro copias de gp16, ocho de gp15 y diez de gp14 (Agirrezabala *et al.*, 2005; Kemp *et al.*, 2005) (Figura 1.8A).

La cola mide 23 nm y tiene un ancho que varía desde 21 nm en la parte proximal hasta 9 nm en el extremo distal, y está formada por gp11 y gp12. A la parte proximal de la cola se unen seis fibras formadas por trímeros de gp17. Cada una de las fibras está formada por un extremo amino terminal que se une a la cola del fago. A continuación, según sugieren las imágenes de microscopía electrónica, hay una región de hélice superenrollada de triple cadena, de unos 120 aminoácidos, que forman la parte proximal de la fibra. La parte distal está formada por dominios globulares (Figura 1.8B). En principio, gp17 se uniría a los LPS a través de su extremo distal (Steven *et al.*, 1988). El mecanismo de unión de las fibras ha sido descrito recientemente: en un primer paso, mientras el fago está más alejado de la superficie celular, las fibras estarían plegadas contra la cápside, de una forma similar a la descrita en otros fagos, lo que permitiría una mayor difusión del fago en el medio (Hu *et al.*, 2013) (Figura 1.8C). Durante la adsorción, las fibras rotan pivotando en su extremo amino terminal, por el que se unen a la cola. La unión de las fibras a la célula es débil y reversible, pero permite al fago acotar la superficie para que la infección proceda eficientemente. Además, la unión cooperativa de hasta seis

fibras podría ser suficientemente específica como para que se llevase a cabo la infección. Alternativamente, un sitio de unión secundario en la cola podría efectuar esta unión específica e irreversible (Hu *et al.*, 2013). Tras la adsorción, se produce una transducción de señal que permite iniciar los pasos irreversibles del proceso de infección (Molineux, 2001). Analizando la capacidad de infección en cepas de *E. coli* K-12 con las proteínas implicadas en la biosíntesis de los LPS mutadas, se ha determinado que la parte más importante de los LPS para la correcta unión de las fibras de T7 es la primera glucosa y la última glucosa con la heptosa (Qimron *et al.*, 2006).

Se ha propuesto que durante la eyección las proteínas del núcleo se despliegan parcialmente para pasar por el canal formado por el conector y la cola, formando una estructura denominada "cola extensible" (Molineux, 2001), en la que gp14 se localiza en la membrana externa mientras que gp15 y gp16 se unen a la membrana interna (Figura 1.8A y C). Para que se pueda formar esta estructura, gp16 presenta actividad transglucosilasa, que permite ampliar el espacio necesario en la membrana para la cola extensible (Molineux, 2001). La extensión de la cola mide unos 40-55 nm de largo y 8-11 de ancho, lo que haría posible atravesar el espacio periplasmático y la membrana citoplasmática, con un canal interno de 3-4 nm para el paso del DNA hacia el interior de la célula (Serwer *et al.*, 2008).

Al abrirse el poro se eyectan de forma eficiente los primeros 850 pb del DNA, donde se encuentra la región que codifica gp0.3, una proteína viral que bloquea las enzimas de restricción de *E. coli* (Hyman & Abedon, 2010), tres promotores para la RNA polimerasa de la bacteria y un promotor de T7. Como gp16 bloquea la translocación del resto del DNA al interior de la célula, es necesaria la actividad de la RNA polimerasa para continuar con el paso del DNA hacia el interior. Mientras que la primera fase de 850 pb se transloca a una velocidad de 75 pb/s, en la segunda fase la RNA polimerasa de *E. coli* internaliza 7 kpb a 40 pb/s y, finalmente, la RNA polimerasa de T7 internaliza el resto del genoma a una velocidad de 200-300 pb/s (Molineux, 2001).

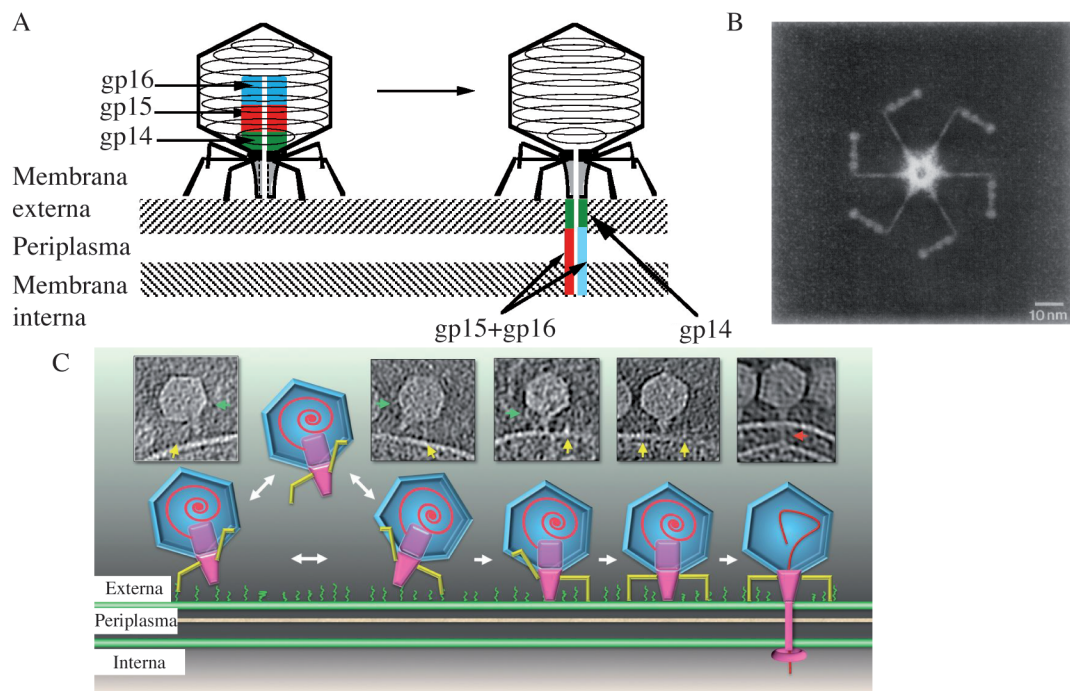


Figura 1.8: El podovirus T7. (A) Representación esquemática del mecanismo de cola extensible (Chang *et al.*, 2010), (B) imagen de microscopía electrónica de las fibras de T7 (Steven *et al.*, 1988), (C) esquema del proceso de infección del fago T7: tras la unión de las fibras el núcleo atraviesa la membrana formando una "cola extensible" que permite la eyección del DNA (Hu *et al.*, 2013).

1.8 Aplicaciones de los bacteriófagos

1.8.1 Terapia fágica

Tras su descubrimiento, Felix d'Herelle empezó a utilizar fagos para el tratamiento de la disentería. Tras realizar los primeros ensayos en animales, pronto empezó a trasladar sus tratamientos a humanos. Para ello, él mismo probó las preparaciones de fagos y al no presentar efectos secundarios, se los administró a varios miembros de su familia determinando así que eran seguros para extender estos tratamientos al resto de la población (Summers, 2001). Sin embargo, tras el entusiasmo inicial y debido al auge de los antibióticos, la terapia de fagos quedó relegada a un segundo plano en Europa occidental y en Estados Unidos, aunque se continuó utilizando en los países de la antigua Unión Soviética y Europa del Este, donde d'Herelle continuaba sus tratamientos. El escepticismo y el abandono de la terapia de fagos durante el apogeo de los antibióticos también fue debido en parte a los deficientes protocolos de obtención de los preparados comerciales, en los cuales

los fagos no estaban purificados correctamente, ya que las muestras comerciales analizadas por d'Herelle y otros científicos en muchas ocasiones no contenían fagos activos. Por ejemplo, en un laboratorio que preparaba cócteles de fagos, Max Delbrück descubrió que ya no se mezclaban stocks independientes, sino que se juntaban antes y se producía una infección de la que se obtenían los fagos para el cóctel, ya que en el laboratorio decían que así se lisaban las bacterias en menos tiempo. Delbrück descubrió que en esta mezcla había un fago que crecía más que los otros y que las muestras finales del supuesto cóctel contenían sólo este fago, que llamó T7 (Summers, 2001). Así pues, aunque algunos científicos intentasen demostrar que la terapia fágica era eficaz, las malas prácticas y la comercialización de antibióticos hicieron que los fagos se redujesen a un modelo bioquímico.

La proliferación de cepas bacterianas resistentes a varios antibióticos ha provocado un nuevo interés por la terapia fágica. Los estudios más recientes relacionados con la terapia de fagos incluyen el desarrollo de cócteles de fagos, de diferentes rutas de distribución, la creación de fagos modificados mediante ingeniería genética con rangos de hospedador ampliados (Marzari *et al.*, 1997) y los fagémidos derivados del fago M13 que expresan proteínas que provocan la muerte celular programada en las bacterias que infectan (Westwater *et al.*, 2003). Otro de los grandes problemas actuales para la eliminación de bacterias patógenas es la presencia de los denominados biofilms, unas capas protectoras de exopolisacáridos (EPS) que dificultan la acción de los antibióticos, pero algunos fagos expresan EPS depolimerasas que destruyen estos biofilms (Cornelissen *et al.*, 2012).

Aunque en los casos anteriormente descritos se utiliza el fago completo, también es posible el uso de proteínas fágicas de forma independiente. Por ejemplo, las proteínas de unión a receptor (RBP), que son más estables que el fago completo y se manipulan con facilidad. Como las RBP determinan la especificidad de hospedador, pueden utilizarse como sensores que aglutinan las bacterias sin producir lisis. Algunas RBP poseen actividad enzimática y degradan los polisacáridos de la envoltura celular, alterando la motilidad de las bacterias (Waseh *et al.*, 2010). Por sí mismas no producen la muerte de la bacteria, pero sí debilitan su envoltura, por lo que constituyen un buen complemento a otros agentes antibacterianos (Meng *et al.*, 2011).

Otras proteínas fágicas de interés son las endolisinas. Estas proteínas degradan el peptidoglicano y, añadidas externamente, provocan la lisis de la bacteria (lisis "desde

fuera"). Hasta el momento sólo se han usado en bacterias Gram positivas ya que la membrana externa de las Gram negativas impide el contacto de las endolisinas con el peptidoglicano. Se han utilizado con éxito frente a *Streptococcus pneumoniae*, *Streptococcus pyogenes* (Fischetti, 2010) y *Staphylococcus aureus* (Fernandes *et al.*, 2012) entre otras.

1.8.2 Phage display

La técnica de *phage display* consiste en el uso de algunos fagos como λ , M13 o T7, que han sido modificados para que expresen en su superficie múltiples copias de diferentes péptidos. Los fagos, una vez modificados, se incuban con su diana previamente inmovilizada para seleccionar uno o varios fagos que se unen eficazmente. Esta técnica se utiliza para crear anticuerpos, para el estudio de interacciones proteína-proteína o para la búsqueda de inhibidores enzimáticos (Pande *et al.*, 2010). Por ejemplo, se ha descrito recientemente el uso del fago M13 para seleccionar péptidos que se unen a la cápsula de *Bacillus anthracis* (Beer y Liu, 2012), o una combinación de *phage display* con otras técnicas bioquímicas para una detección más sensible del virus de la hepatitis (Monjezi *et al.*, 2013).

1.8.3 Tipado bacteriano

Esta técnica, empleada de forma individual o en combinación con el serotipado tradicional, consiste en incubar una batería de fagos con las bacterias desconocidas y el posterior análisis del grado de lisis obtenido, definiendo como tipo fágico al conjunto de aislados bacterianos que presentan un mismo patrón de reacción frente a diferentes fagos (Frost *et al.*, 1999). Los primeros métodos estandarizados se emplearon para tipar *Staphylococcus* en la década de los 60 (Blair y Williams, 1961). Posteriormente, en los 70, se describió el primer sistema para tipar *Listeria monocytogenes*. Además de permitir el tipado de la bacteria, estos ensayos se han utilizado con fines epidemiológicos, y gracias a su gran especificidad, han permitido localizar el origen del brote (Audurier y Martin, 1989). También se estandarizaron los sistemas de tipado para *Campylobacter jejuni* y *Campylobacter coli* (Grajewski *et al.*, 1985) entre otros. Estos sistemas se han mantenido en auge hasta la aparición de las técnicas de secuenciación masiva que desplazaron su uso.

1.8.4 *Reporter phages*

En ocasiones, establecer con la mayor celeridad posible la presencia de una bacteria en una muestra es trascendental para poder aplicar un tratamiento eficaz contra algunas enfermedades. A veces, el crecimiento de un cultivo de la bacteria que provoca la enfermedad para poder identificarla puede llevar días. En estos casos el uso de fagos específicos para estas bacterias se presenta como una alternativa más rápida. En este caso, más que eliminar a las bacterias patógenas, lo que se pretende es su identificación, utilizando el fago como mero indicador de presencia.

A modo de ejemplo se puede describir la detección de la peste, que es una enfermedad que requiere una detección rápida, pero el ciclo de *Yersinia pestis* es lento y se necesitan más de dos días para obtener colonias de la bacteria. El fago ϕ A1122, similar a T7, se ha modificado para expresar una proteína bioluminiscente, de forma que a la vez que se multiplica en el interior de la célula, se genera una respuesta bioluminiscente, reduciendo el tiempo de detección de dos días a dos horas (Schofield *et al.*, 2012). Este tipo de "fluorofagos" también se ha utilizado para la detección de *Mycobacterium tuberculosis*, una bacteria de crecimiento muy lento, reduciendo el tiempo de detección de 4-8 8 semanas a menos de dos días (Jain *et al.*, 2012).

1.9 *Conclusión*

Los bacteriófagos han sido ampliamente estudiados desde el punto de vista bioquímico desde su descubrimiento, llegando a convertirse en sistemas modelos de muchos procesos biológicos esenciales. Las técnicas modernas complementan los trabajos que habían llevado a cabo d'Herelle y sus seguidores, lo que permite una mejor comprensión de los mismos y una actualización de las aplicaciones que ya han tenido desde su descubrimiento.

Uno de los aspectos menos caracterizados del ciclo de replicación del fago es la unión a la célula. Los fagos que poseen fibras realizan un primer reconocimiento de la superficie de la célula con ellas, determinando si una célula es adecuada y uniéndose de forma reversible hasta localizar la mejor posición para la infección. Aunque en muchos casos las fibras no son imprescindibles, sí mejoran la eficacia de la infección, lo que puede ser vital en el medio natural. Un análisis exhaustivo de este proceso, permitiría introducir modificaciones que alteren el rango de hospedadores, optimizando la manipulación de fagos en terapias futuras.

Durante las últimas dos décadas se han incrementado notablemente los estudios estructurales de los bacteriófagos, principalmente mediante (crio)-microscopía electrónica y más recientemente mediante cristalografía de rayos X. En los estudios por microscopía electrónica de los fagos completos, donde las fibras son visibles, estas no se localizan en las reconstrucciones tridimensionales de las partículas virales porque son estructuras muy flexibles que no siguen las simetrías icosaédrica y helicoidal de la cabeza y la cola, respectivamente. Desde un punto de vista cristalográfico, los esfuerzos han estado centrados en la cristalización de espigas, probablemente debido a la presencia de actividad hidrolítica.

Al comenzar las investigaciones correspondientes esta tesis doctoral no había disponible información estructural de alta resolución de fibras de fagos, aunque se había estudiado en detalle el ensamblaje y función de las de T7 y T4 (Steven *et al.*, 1988; Cerritelli *et al.*, 1996; Leiman *et al.*, 2003). Recientemente se ha descrito la estructura atómica de las fibras largas del myovirus T4 (Bartual *et al.*, 2010b), además de haberse estudiado en detalle su ensamblaje y función (Leiman *et al.*, 2003). En esta tesis doctoral se presentan las estructuras atómicas de las fibras del podovirus T7 y del siphovirus T5.

2 Objetivos

El principal objetivo de esta tesis doctoral es la caracterización estructural de las fibras de los bacteriófagos T7 y T5. Para ello, se han establecido los siguientes objetivos específicos:

1. Clonación y expresión de la fibra del bacteriófago T7 formada por la proteína gp17, selección de un fragmento carboxilo terminal estable para su cristalización y resolución de su estructura.
2. Caracterización de la fibra del bacteriófago T5 formada por la proteína pb1, selección de un fragmento carboxilo terminal estable para su cristalización y resolución de su estructura.
3. Análisis de la estructura atómica de las proteínas gp17 y pb1, y comparación con otras proteínas.

3 Materiales y métodos

3.1 Métodos biológicos

3.1.1 Cepas bacterianas

A continuación se enumeran las cepas bacterianas utilizadas en este trabajo para las clonaciones y las expresiones.

- XL1-Blue: cepa de *E. coli*: *recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F' *proAB lacI^qZΔM15 Tn10* (Tet^r)] (Stratagene).
- BL21(DE3): cepa de expresión de *E. coli* B: F⁻, *ompT*, *hsdS_B*(r_B⁻ m_B⁻) *dcm*, *gal*, *lon*, λ (DE3) (Promega).
- JM109(DE3): cepa de expresión de *E. coli*: *endA1, recA1, gyrA96, thi, hsdR17*(r_k⁻, m_k⁺), *relA1, supE44*, Δ (*lac-proAB*), (F' *traD36, proAB, lacI^qZ ΔM15*), λ (DE3) (Promega).
- BL21(DE3)pLysS: cepa de expresión de *E. coli* B: F⁻, *ompT*, *hsdS_B*(r_B⁻, m_B⁻), *dcm*, *gal*, *lon*, λ (DE3), pLysS, Cm^r (Promega).
- DH5 α : cepa de *E. coli*: F⁻, Φ 80*lacZ ΔM15*, Δ (*lacZYA-argF*), U169, *recA1, endA1, hsdR17*, (rK⁻, mK⁺), *phoA supE44 λ-thi-1 gyrA96 relA1* (Invitrogen).
- CR63: cepa permisiva de *E. coli*: F⁺, λ^- , *serU60*(AS), *lamB63*
- F: cepa de *E. coli* de adsorción rápida. Descrita por Lanni (Lanni, 1958)
- BL21: cepa de *E. coli* B: F⁻, *ompT*, *hsdS_B*(r_B⁻ m_B⁻) *dcm*, *gal*, *lon* (Promega).
- Rosetta(DE3)pLysS: cepa de expresión de *E. coli* B: F⁻, *ompT*, *hsdS_B*(r_B⁻ m_B⁻), *gal*, *dcm*, *lon*, λ (DE3), pLysSRARE (Cam^R) (Novagen).
- Rosetta2(DE3)pLysS: cepa de expresión de *E. coli* B: F⁻, *ompT*, *hsdS_B*(r_B⁻ m_B⁻), *gal*, *dcm*, *lon*, λ (DE3), pLysSRARE2 (Cam^R) (Novagen).
- B834(DE3): cepa auxotrofa para metionina de *E. coli* B: F⁻, *ompT*, *hsdS_B*(r_B⁻ m_B⁻), *gal*, *dcm*, *met*, λ (DE3) (Novagen).

Para la expresión de proteínas se utilizan cepas que contienen el lisógeno λ (DE3), que incluye el gen *lacI*, el promotor *lacUV5* y el gen para la RNA polimerasa de T7. El gen *lacI* es un represor del promotor *lacUV5*, que provoca que en ausencia

de IPTG (Isopropil β -D-tiogalactósido) no se produzca la transcripción de la RNA polimerasa. Cuando se induce la expresión, el IPTG bloquea al represor, y el promotor queda libre permitiendo la expresión de la RNA polimerasa y por lo tanto de la proteína recombinante codificada en el vector de expresión.

Todas las cepas derivadas de *E. coli* B carecen de la proteasa *lon* y de la proteasa *ompT* de la membrana externa. Estas proteasas podrían degradar las proteínas durante la expresión, por lo que en muchas ocasiones las proteínas expresadas son más estables en estas cepas. Por otra parte, la cepa JM109(DE3) deriva de la cepa K-12 y sí presenta estas proteasas, pero su crecimiento es más lento, por lo que facilita el plegamiento de algunas proteínas recombinantes.

Hay que tener en cuenta que incluso antes de la inducción se puede producir expresión basal de la RNA polimerasa de T7 y, por lo tanto, de la proteína codificada por el vector insertado. Por ello, algunas de las cepas utilizadas contienen el plásmido pLysS. Este plásmido proporciona a la célula pequeñas cantidades de lisozima de T7, que actúa como inhibidor de la RNA polimerasa de T7, inhibiendo la transcripción, lo que estabiliza los plásmidos que expresan proteínas que pueden ser tóxicas.

También se han utilizado dos cepas Rosetta. Estas cepas tienen un plásmido que además de expresar la lisozima de T7 expresan genes de tRNA raros en *E. coli*, lo que evita que la expresión dependa de la disponibilidad de estos codones en la bacteria. En el vector pLysSRARE se codifican los tRNA para AUA, AGG, AGA, CUA, CCC y GGA. En el caso del vector pLysSRARE2, además se codifica el CGG.

3.1.2 Plásmidos

- pCR2.1-TOPO (Invitrogen)
- pET28a(+) (Novagen)
- pET28c(+) (Novagen)
- pET30a(+) (Novagen)
- pETDuet (Novagen)

Para la expresión de las proteínas se han utilizado vectores del sistema pET. Estos vectores además del promotor *lacUV5* presente en las células, están controlados por el promotor *T7lac*, que se encuentra en el propio vector. De esta forma, el represor *lac* actúa simultáneamente en la célula, evitando la expresión de la RNA polimerasa,

y en el vector, de forma que aunque exista expresión basal de la RNA polimerasa, ésta no pueda unirse al promotor en el vector (Studier *et al.*, 1990).

Los vectores pET28 y pET30 son vectores con resistencia a kanamicina y con un único sitio de clonación. Ambos permiten clonar el inserto dejando colas de histidinas en los extremos amino o carboxilo terminal. El pET28 incluye además una cola T7 y el pET30 una cola S. En ambos vectores la cola de histidinas se puede eliminar tras la purificación cortando con trombina y en el caso del pET30 también se puede cortar la cola S con enterokinasa (Studier *et al.*, 1990). En el caso del pET28 se han utilizado las variantes a+ y c+, que se diferencian en su pauta de lectura, existiendo dos nucleótidos extra en el caso del c+.

El vector pETDuet es similar a los anteriores, pero tiene dos sitios de clonación, lo que permite expresar dos proteínas simultáneamente, y resistencia a ampicilina. Los vectores del sistema Duet se eligen en los casos en los que existe la posibilidad de necesitar coexpresar dos proteínas, eliminando la necesidad de incluir un segundo vector en la célula (Novy *et al.*, 2002).

3.1.3 Medios de cultivo

- LB (Luria-Bertani): 10 g/l triptona, 5 g/l extracto de levadura, 10 g/l cloruro sódico.
- LB-agar: 10 g/l triptona, 5 g/l extracto de levadura, 10 g/l cloruro sódico, 15 g/l bacto-agar.
- LB-topagar: 10 g/l triptona, 5 g/l extracto de levadura, 10 g/l cloruro sódico, 7,5 g/l bacto-agar.
- TSS: basado en el descrito por Chung (Chung *et al.*, 1989): LB con 12% (m/v) PEG 8000, 5% (v/v) DMSO, 25 mM cloruro magnésico y 1% (v/v) glicerol.

El medio de cultivo LB es el más utilizado para crecimiento de bacterias en medio líquido. Paralelamente, para los cultivos sólidos en placas Petri, se emplea el mismo medio suplementado con un 1,5% (m/v) de bacto-agar si las células son extendidas sobre la superficie del medio y con un 0,7% si son crecidas en el interior del medio semisólido.

El medio TSS se ha utilizado para hacer células competentes. El cloruro magnésico proporciona a las células la competencia química, incrementada por el resto de elementos del medio, y la presencia de PEG y el glicerol facilita el

almacenamiento de las células a -80 °C sin que sufran daños durante el proceso de congelación.

3.2 Manipulación de células bacterianas

3.2.1 Obtención de células competentes

Para la obtención de células competentes se creció la cepa bacteriana en medio Luria-Bertani (LB) a 37 °C toda la noche. A la mañana siguiente se realizó una dilución 1:100 en el mismo medio y se creció hasta alcanzar una DO_{600} (densidad óptica a 600 nm) de 0,3. Se centrifugaron a 3000 x g durante 10 min a 4 °C y se resuspendieron en 10:1 de su volumen en medio TSS, guardándose en alícuotas de 250 µl hasta su uso a -80 °C.

3.2.2 Transformación química bacteriana

Las células competentes se descongelaron en hielo y se mezcló una alícuota de 50 µl de bacterias con 50 ng del plásmido. Se incubó en hielo durante 30 min y se realizó un choque térmico a 42 °C durante 45 s. Se incubó en hielo durante 2 min y se añadieron 200 µl de medio LB. Se realizó una incubación de 1 hora a 37 °C con agitación y se sembraron en una placa con LB-agar suplementada con el antibiótico correspondiente: la ampicilina se suplementó a una concentración final de 100 µg/ml y la kanamicina y estreptomycin a 50 µg/ml.

3.3 Manipulación de ácidos nucleicos

3.3.1 Amplificación del ácido nucleico

Las reacciones se realizaron con el kit *Taq DNA Polymerase* (New England Biolabs) siguiendo las instrucciones del fabricante. En resumen, para cada reacción se añadieron 0,2 µM de cada uno de los oligonucleótidos y 0,2 mM de dNTPs (desoxirribonucleótidos). La PCR se realizó con una desnaturalización inicial de 4 min a 94 °C, seguida de 35 ciclos de desnaturalización de 30 s, hibridación de 1 minuto a la temperatura indicada para los oligonucleótidos, y amplificación a 72 °C de 1 minuto/kb, seguidos de una amplificación final de 10 min. Los oligonucleótidos empleados en cada caso se indican en el Anexo 1.

Los productos amplificados fueron purificados con el kit *QIAquick PCR Purification* (Qiagen).

3.3.2 Corte con enzimas de restricción

Las reacciones con enzimas de restricción se llevaron a cabo según las instrucciones del proveedor (Promega, Thermo Fisher o New England Biolabs) con los tampones suministrados y a 37 °C.

3.3.3 Aislamiento de fragmentos de DNA y purificación

Para la purificación de fragmentos obtenidos como producto de las restricciones enzimáticas se realizaron electroforesis en geles de agarosa al 1% (m/v) en TAE (40 mM Tris-HCl, 20 mM ácido acético, 1 mM EDTA, pH 8,0) y 0,5 µg/ml de bromuro de etidio. Para cargar las muestras se mezclaron con tampón EA 6x (0,25% (m/v) azul de bromofenol, 0,25% (m/v) xilenocianol, 15% (v/v) Ficoll). Se cortaron las bandas de interés con una cuchilla y se purificaron con el kit *GeneJET Gel Extraction* (Thermo Fisher).

3.3.4 Purificación del DNA plasmídico

La purificación del DNA plasmídico se realizó a partir de 20 ml de cultivo bacteriano saturado procedente de una colonia aislada, utilizando los kits *Wizard Plus SV Minipreps* (Promega) o *GeneJet Plasmid Miniprep* (Thermo Fisher).

3.3.5 Ligación y obtención de vectores

Las reacciones de ligación se realizaron durante 4 h a 20 °C o durante 16 h a 4 °C utilizando la DNA ligasa del fago T4 (Promega), con 50 µg de vector y manteniendo una proporción vector-inserto de 1:3, según las instrucciones del fabricante. Los productos de ligación fueron transformados posteriormente en células competentes XL1-Blue (en el caso de las proteínas de T7) o DH5α (en el caso de las proteínas de T5). La presencia de inserto en las colonias obtenidas se comprobó en un primer paso mediante PCR con los oligonucleótidos utilizados para la amplificación del fragmento en el caso del vector pETDuet o pCR2.1-TOPO y los oligonucleótidos *T7 promoter primer* y *T7 terminator primer* para el resto de vectores. Una vez aislado el vector, se comprobó la presencia del inserto por restricción y se secuenció para comprobar la ausencia de mutaciones y la correcta pauta de lectura (Sistemas Genómicos SL o Secugen SL).

3.4 Manipulación de proteínas

3.4.1 Análisis por SDS-PAGE

Las electroforesis en geles de poliacrilamida se llevaron a cabo según el protocolo de Laemmli (Laemmli, 1970). La concentración de poliacrilamida en el gel separador es del 12,5% (m/v) (excepto en los casos en los que se indique lo contrario) y la del concentrador del 4%. El tampón de electroforesis era 25 mM Tris-HCl, pH 8,3, 250 mM glicina, 0,1% (m/v) SDS y el tampón de carga utilizado fue EP 6x (100 mM Tris-HCl pH 6,8, 4% (m/v) SDS, 200 mM DTT, 20% (v/v) glicerol y 0,2% (m/v) azul de bromofenol). Tras la electroforesis, las proteínas se tiñeron con azul de Coomassie al 0,25% (m/v) en 33% (v/v) etanol y 10% (v/v) ácido acético y se destiñeron en la misma solución sin colorante.

En gran parte de los casos se analiza la presencia de la proteína por duplicado, en una fracción hervida y otra sin hervir, ya que las fibras forman trímeros que en muchos casos son resistentes al SDS. Esta doble confirmación permite saber que además de tener el peso molecular correcto, si la proteína forma un trímero.

3.4.2 Análisis por Tricina-SDS-PAGE

Las electroforesis en geles de tricina se llevó a cabo según el protocolo de Schagger (Schagger, 2006). La concentración de poliacrilamida en el separador es del 16% (m/v) y en el concentrador del 4%. La tinción y desteñido se realizó del mismo modo que con los geles SDS-PAGE.

3.5 Purificación mediante cromatografía

A continuación se detallan las diferentes técnicas cromatográficas utilizadas en el desarrollo de este trabajo. Las cromatografías de afinidad se han hecho siempre por gravedad, mientras que para las cromatografías de intercambio iónico y de exclusión molecular se ha utilizado un FPLC modelo *BioLogic DuoFlow* (Bio-Rad) o *ÄKTA Purifier* (GE Healthcare).

3.5.1 Cromatografía de afinidad

La cromatografía de afinidad separa las proteínas basándose en una interacción reversible de la proteína con un ligando específico unido a la resina. Esta técnica ofrece una gran selectividad y, por lo tanto, resolución. Para eluir la proteína de la columna la interacción debe ser revertida, ya sea específicamente utilizando un

ligando que compita con la proteína o la resina, o inespecíficamente cambiando el pH, la fuerza iónica, la polaridad o añadiendo EDTA (Figura 3.1A).

Las proteínas clonadas con una cola de seis histidinas en el extremo amino terminal fueron purificadas utilizando una resina de níquel (Jena Biosciences), que tiene una elevada afinidad por esta cola de histidinas. Las proteínas se eluyeron mediante una interacción competitiva, utilizando concentraciones crecientes de imidazol.

3.5.2 Cromatografía de intercambio iónico

La cromatografía de intercambio iónico separa las proteínas basándose en las diferencias de carga en la superficie de las proteínas. Se basa en la formación de interacciones iónicas entre los grupos polares cargados de la superficie de las proteínas y la resina de la columna. Las proteínas con zonas de carga en la superficie opuestas a las de la resina al pH empleando son retenidas, mientras que el resto no interaccionan y son eluidas. Las proteínas que han quedado retenidas, se eluyen posteriormente con un tampón con mayor fuerza iónica o con un pH diferente (Figura 3.1C).

En este trabajo se describen purificaciones con columnas de intercambio aniónico (Q) y catiónico (S). En ambos casos se realizó la elución con un gradiente en el mismo tampón, aumentando gradualmente la concentración de cloruro sódico.

3.5.3 Cromatografía de exclusión molecular

La cromatografía de exclusión molecular separa las proteínas basándose en las diferencias de tamaño. La resina de las columnas está compuesta por partículas esféricas con poros de un tamaño determinado, de forma que las proteínas difunden a una velocidad proporcional a su tamaño. Como las proteínas en principio no interaccionan con la resina, esta cromatografía es relativamente independiente del tampón utilizado (Figura 3.1B).

3.5.4 Cromatografía de interacción hidrofóbica

La cromatografía de interacción hidrofóbica separa las proteínas basándose en su hidrofobicidad. Las proteínas que están en un tampón de gran fuerza iónica interaccionan con la resina de la columna y quedan retenidas. La elución se realiza disminuyendo la fuerza iónica del tampón, al disminuir la concentración de la sal utilizada, que habitualmente es sulfato amónico.

En este trabajo se describe una purificación con una columna de interacción hidrofóbica de fenil sefarosa. El gradiente decreciente de sal se realizó con sulfato amónico.

3.5.5 Cromatografía en columna de hidroxiapatito

La cromatografía de hidroxiapatito separa las proteínas según la interacción de la proteína con la resina de la columna. La resina está formada por fosfato de calcio, por lo que posee cargas tanto positivas como negativas y la unión de la proteína depende de la interacción con estas cargas, que provoca la adsorción a la resina. Para la elución se aumenta la concentración de fosfato sódico.

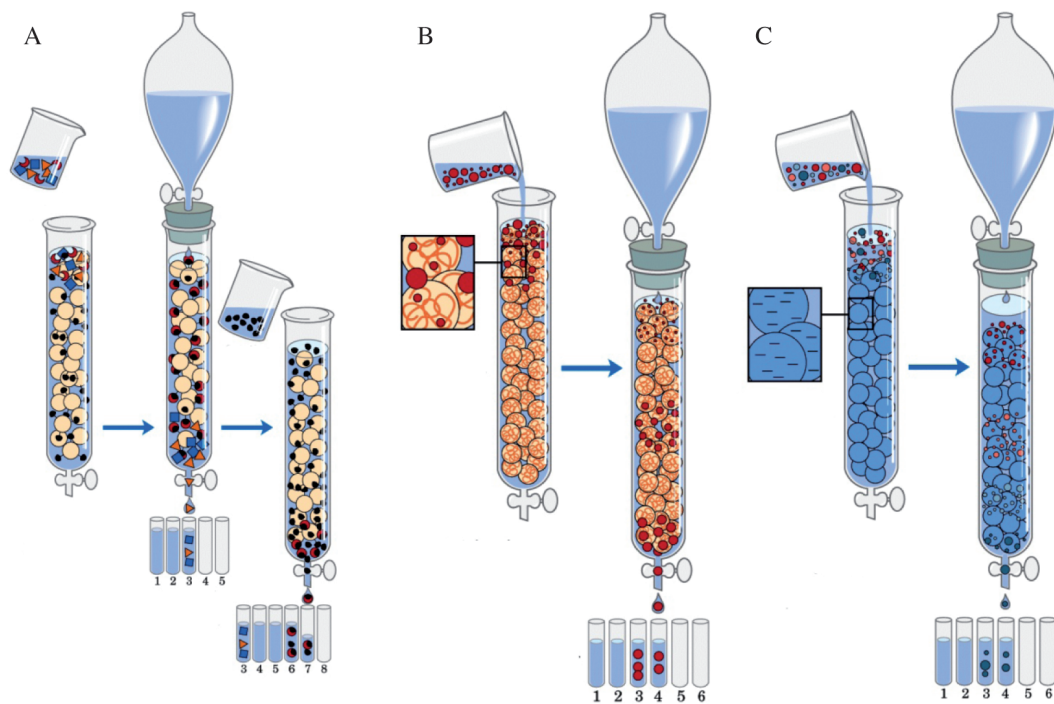


Figura 3.1: Principales tipos de técnicas cromatográficas. (A) Afinidad: la proteína de interés, en rojo, interacciona con la columna hasta que se añade un competidor, en negro, que hace que la proteína se eluya. (B) Exclusión molecular: las proteínas eluyen en función de su tamaño. (C) Intercambio iónico: las proteínas eluyen en función de su carga en un gradiente iónico, la diferencia de cargas se representa con diferentes colores.

3.6 Técnicas cristalográficas

3.6.1 Naturaleza de los cristales de proteína

Para que una proteína pueda cristalizar, conviene que las muestras cumplan una serie de características como una alta homogeneidad, alta solubilidad, alta estabilidad

y un alto grado de pureza (Bergfors, 2009). Las muestras que cumplan estos requisitos se utilizan para cristalizar, ya que así aumenta la probabilidad de que den lugar a cristales que difracten los rayos X. Además, para que la muestra pueda cristalizar, la proteína debe estar supersaturada en la gota.

El proceso de cristalización se puede dividir en dos pasos: nucleación y crecimiento. Durante la nucleación, las moléculas pasan de un estado desordenado a uno ordenado que permite formar los primeros agregados paracristalinos, denominados núcleos. Una vez que el núcleo tiene un tamaño determinado comienza la fase de crecimiento, que continúa mientras no se acumulen defectos en la superficie del cristal que impidan su aumento de tamaño y mientras haya supersaturación de la proteína en la solución (McPherson, 2004).

La celdilla unidad se define por tres ángulos y tres ejes, y se obtiene mediante la aplicación de operaciones de simetría a la unidad mínima de repetición, que es la unidad asimétrica. Los cristales de proteína son, por lo tanto, redes ordenadas en las que un elemento (celdilla unidad) se repite en las tres direcciones. Las operaciones de simetría definen los 230 grupos espaciales posibles, que se limitan a 65 en el caso de las proteínas debido a su quiralidad (McPherson, 2004).

Los cristales de proteína suelen ser pequeños, pero presentando celdillas unidad de gran tamaño. En el caso de los cristales de fibras, uno de los ejes tiende a ser especialmente largo. Además, los cristales de proteína suelen tener un alto contenido de solvente permite el correcto plegamiento, pero que incrementa la fragilidad del cristal, que sumada a su tamaño hace que los cristales sean susceptibles de ser dañados al manipularlos, por la radiación o por variaciones térmicas (McPherson, 2004).

3.6.2 Proceso de cristalización

Para que se pueda llevar a cabo la cristalización, es necesario obtener una solución sobresaturada de proteína, lo que se consigue mezclando la proteína con una solución precipitante. Los precipitantes más usados son polímeros orgánicos, solventes orgánicos o sales. Además, es importante mantener un pH constante, por lo que a la solución (o condición) de cristalización se le añade un tampón.

Existen diferentes técnicas para hacer que la proteína entre en contacto con la solución precipitante y se produzca la cristalización. Las principales son la difusión de vapor, el *microbatch*, la diálisis y la contradifusión.

El método más utilizado, y empleado también en esta tesis, es la difusión de vapor (Figura 3.2). En esta técnica una pequeña gota de proteína se mezcla con una gota de precipitante y el conjunto se equilibra contra un reservorio en el que hay un volumen mucho mayor de precipitante. Durante el proceso de equilibrado, se produce una difusión de vapor de agua de la gota hacia el reservorio, lo que provoca que la proteína quede sobresaturada y, si las condiciones son idóneas, pueda empezar la nucleación. Este método presenta dos variantes: la gota sentada y la gota colgante. En el primer caso la gota de proteína se deposita sobre un soporte (Figura 3.2A) y en el segundo la gota se sitúa sobre un cubreobjetos que queda suspendido sobre el reservorio (Figura 3.2B) (Bergfors, 2009).

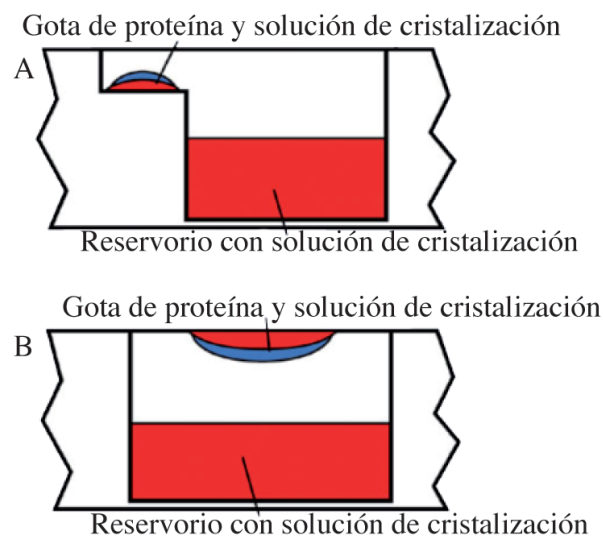


Figura 3.2: Técnicas empleadas para llevar a cabo el proceso de cristalización. (A) Difusión de vapor en gota sentada. (B) Difusión de vapor en gota colgante. Imagen adaptada (Rupp, 2009).

Para encontrar la condición idónea de cristalización se realiza una búsqueda inicial y un proceso de optimización posterior. La búsqueda inicial puede llevarse a cabo con un robot de cristalización, permitiendo utilizar una menor cantidad de proteína, lo que aumenta el número de condiciones diferentes que se pueden probar. Inicialmente, las pruebas se llevan a cabo utilizando soluciones comerciales que hacen un cribado aleatorio de decenas o cientos de condiciones de cristalización diferentes entre sí. A continuación se lleva a cabo la optimización en la que se escala el tamaño de la gota si la búsqueda inicial se hizo con robot, y se ajusta el pH, la concentración de precipitante, la concentración de proteína y la temperatura a la que

se hace el experimento. También se pueden emplear aditivos como alcoholes, sales o detergentes, que en algunos casos mejoran la cristalización (Bergfors, 2009).

3.6.3 Recogida de datos

Conviene medir los cristales obtenidos a 100 K para una mayor estabilidad del cristal frente a la radiación por los rayos X, pero para ello es necesario crioproteger los cristales. La crioprotección consiste en la vitrificación del solvente a baja temperatura con un agente crioprotector. El glicerol, el PEG, o algunas sales a alta concentración son buenos crioprotectores y, si no están presentes en la condición de cristalización, deben añadirse antes de sumergir el cristal en nitrógeno líquido para su congelación (Garman y Owen, 2007), para evitar generar hielo cristalino, que daría lugar a la presencia de anillos de hielo en las imágenes tomadas.

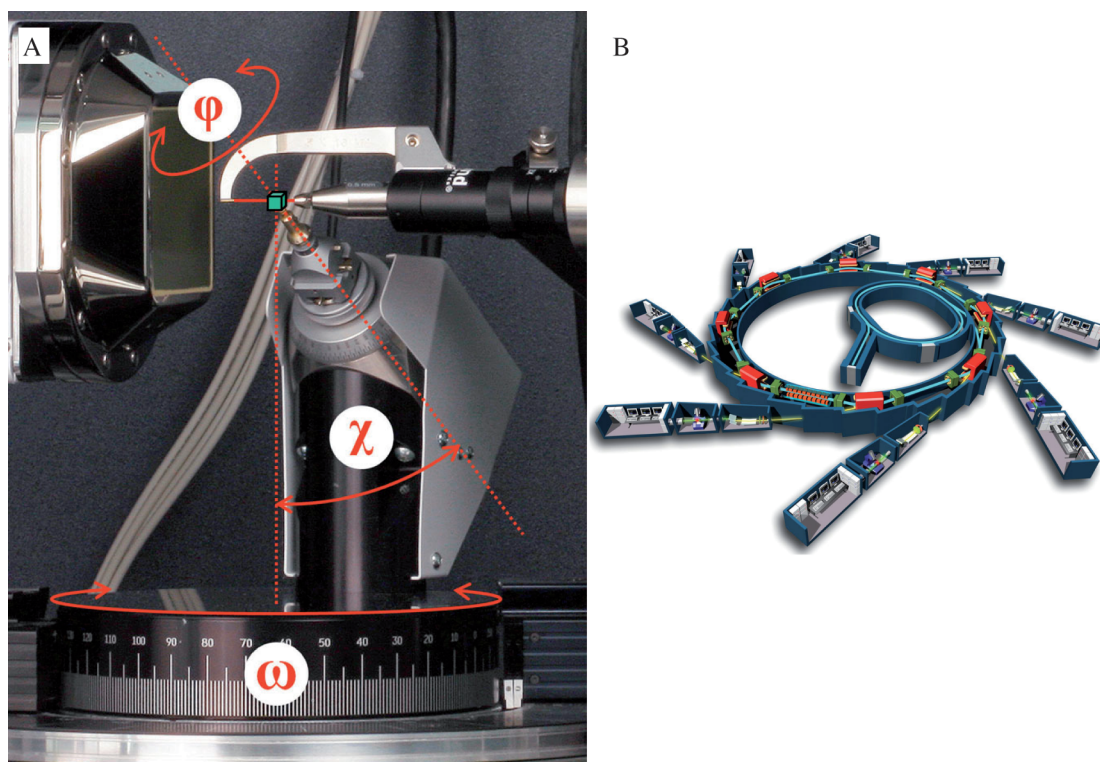


Figura 3.3: La recogida de datos de difracción. (A) Cristal montado en un difractómetro. Se señalan los diferentes ángulos de giro que permiten cambiar la posición del cristal. (B) Esquema de un sincrotrón. Imágenes adaptadas (Rupp, 2009).

Para la recogida de datos se expone el cristal a un haz monocromático de rayos X (Figura 3.3A). Las imágenes son tomadas con un tiempo de exposición determinado mientras el cristal está girando, por lo que en cada imagen se recogen datos correspondientes a un sector del cristal (método de oscilación), hasta obtener un set

de datos completo. La radiación sincrotrónica se genera en aceleradores de partículas con forma de anillo (Figura 3.3B). Frente a las fuentes tradicionales de rayos X, la radiación sincrotrónica tiene más brillo y permite seleccionar la longitud de onda, lo que permite utilizar cristales más pequeños y utilizar métodos de dispersión anómala para resolver el problema de la fase (Moffat y Ren, 1997).

De las imágenes obtenidas se extraen las distintas reflexiones, de las que se obtiene información sobre los parámetros de la celdilla unidad y de la simetría del cristal; y al medir la intensidad de cada una de las reflexiones se obtiene información del contenido de la celdilla y del grupo espacial.

3.7 Resolución de la estructura

3.7.1 El problema de la fase

Para resolver la estructura de la proteína es necesario conocer la densidad electrónica en cada punto de la celdilla unidad. La densidad electrónica se relaciona con los factores de estructura mediante la transformada de Fourier (Figura 3.4). Como los factores de estructura son ondas, se componen de un módulo (las intensidades obtenidas experimentalmente) y una fase, que se pierde al realizar el experimento, dando lugar al problema de la fase (Figura 3.4). Para resolver este problema existen varios métodos:

- Métodos directos: se emplean en la resolución de moléculas pequeñas y en macromoléculas para conocer las subestructuras de los átomos pesados.
- Métodos basados en la introducción de átomos pesados:
 - Reemplazamiento isomorfo simple o múltiple (SIR/MIR): se introducen átomos pesados sin que alteren la estructura cristalina, con el objetivo de utilizar su posición para calcular las fases.
 - Dispersión anómala simple o múltiple (SAD/MAD): se utilizan átomos pesados que sean buenos dispersores y para localizar su posición se usa la diferencia en la dispersión a diferentes longitudes de onda (MAD) o a una única longitud de onda comparando las parejas de reflexiones simétricas (SAD).
- Método de reemplazamiento molecular (MR): para calcular las fases se emplean las fases de una proteína modelo.

Los métodos utilizados en este trabajo han sido el de dispersión anómala (simple y múltiple), los métodos directos (para la localización de los átomos pesados) y el reemplazamiento molecular.

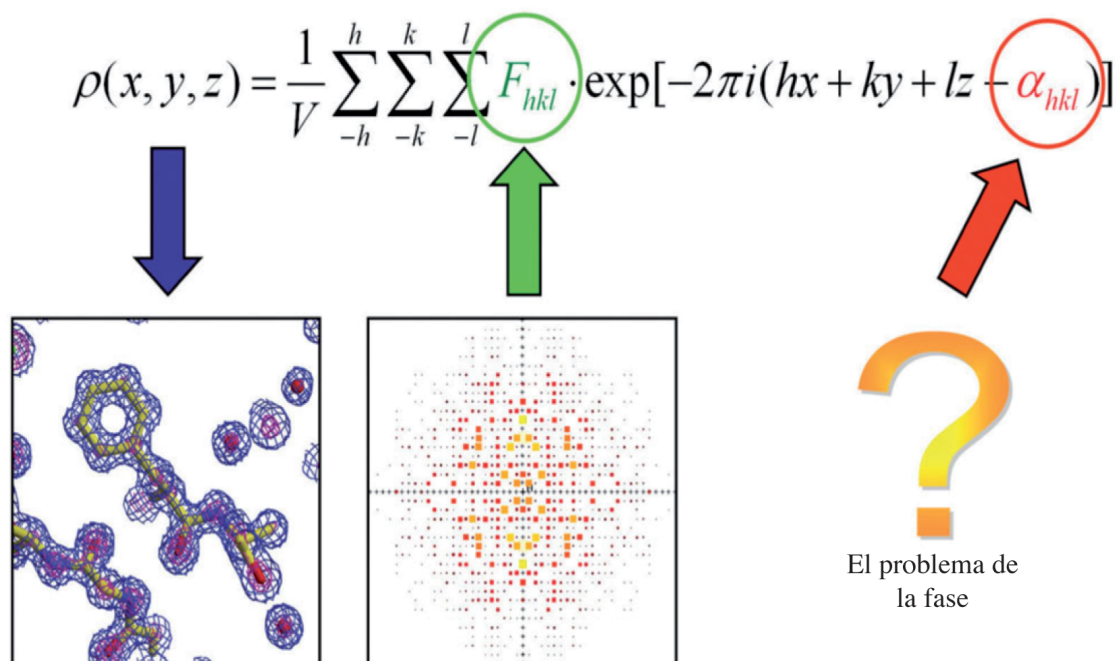


Figura 3.4: El problema de la fase. La densidad electrónica se relaciona con las amplitudes medidas mediante la transformada de Fourier. Al realizar el experimento se pueden medir los factores de estructura (en verde), pero no se puede medir la fase de la onda (en rojo), y generando el problema de la fase. Para poder calcular la densidad electrónica, es necesario resolver este problema. Imagen adaptada (Rupp, 2009).

3.7.2 Dispersión anómala simple/múltiple

La dispersión anómala es la producida por los átomos cuando absorben la radiación, y se produce con pequeñas variaciones en la amplitud y la fase. Este efecto se puede medir al comparar parejas de reflexiones que sean simétricas (pares de Friedel) que deberían ser iguales en intensidad pero opuestas en el signo de la fase. Las diferencias que se observan son las diferencias anómalas o de Bijvoet, y se pueden utilizar para localizar los dispersores anómalos (Walsh *et al.*, 1999).

Para poder obtener una señal anómala que permita calcular las fases, se introducen en la estructura átomos pesados, ya que son mejores dispersores. Los átomos se pueden añadir mediante cocrystalización, inmersión o expresando de proteínas con selenio en lugar de azufre en el medio de crecimiento (Hendrickson *et al.*, 1990).

Para que estas diferencias sean evidentes, se recogen datos al menos a tres longitudes de onda diferentes. Una de ellas ha de ser en la que se maximicen los efectos de la absorción (pico), otra en la que se minimice la dispersión (punto de inflexión) y al menos una en la que se produzca poca dispersión anómala (remoto) (Walsh *et al.*, 1999; Rupp, 2009). En el caso de un experimento de dispersión anómala simple, se recoge una sola longitud de onda, la correspondiente al pico, pero se maximiza la multiplicidad de los datos para poder localizar las diferencias anómalas.

3.7.3 Reemplazamiento molecular

Para esta técnica se emplea como modelo una proteína que se presupone que es suficientemente similar en su estructura a la proteína problema. Se calcula una función de rotación y una de traslación para el modelo de forma que se sitúa sobre la celdilla de la proteína problema, permitiendo calcular las fases. Estas fases iniciales y las amplitudes experimentales se utilizan para calcular la densidad electrónica (Rossmann y Blow, 1962; Rossmann, 1990). Tras obtener una densidad electrónica inicial, se emplean métodos que permiten mejorar la calidad de los mapas como el aplanado de solvente o el promediado por simetría no cristalográfica.

3.7.4 Construcción del modelo

Una vez obtenidas unas fases suficientemente buenas, el modelo inicial se puede construir en muchas ocasiones con programas de forma automática y posteriormente se realiza un ajuste manual a la densidad electrónica completando el modelo inicial. Este modelo tiene que ser refinado para optimizar la relación entre los datos observados y los calculados, modificando las coordenadas de los átomos y el factor térmico (relativo a la movilidad de cada átomo) (Rupp, 2009).

3.7.5 Validación

En un último paso se lleva a cabo una validación de la estructura de forma objetiva. Para ello se comprueba que no haya habido errores, especialmente durante el refinamiento de la estructura. Se comprueba la ausencia de datos atípicos como los ángulos improbables de la cadena principal utilizando el diagrama de Ramachandran (Ramachandran *et al.*, 1963), los de las cadenas laterales, los rotámeros y los contactos entre todos los átomos de la estructura. También se tienen en cuenta los factores térmicos de cada aminoácido, ya que una desviación respecto a la media

puede indicar la presencia de un error en el modelo en una región determinada. Tras haber calculado estas estadísticas se puede volver al paso anterior y manualmente corregir los errores que se hayan detectado durante la validación, hasta obtener el modelo con las mejores estadísticas posibles. La estructura validada, junto con los factores de estructura, se deposita en el Banco de Datos de Proteínas (PDB), en el que antes de poder hacerse pública vuelve a ser comprobada y se le asigna un código (Kleywegt, 2000). Posteriormente esta estructura será liberada, de forma que cualquier usuario pueda descargar las coordenadas y los factores de estructura de la proteína, para poder analizar tanto la estructura como su mapa de densidad.

3.8 La proteína gp17

3.8.1 Clonación del gen de gp17

El DNA del bacteriófago T7 fue proporcionado por la Dra. Ana Cuervo, el Prof. José L. Carrascosa y el Dr. Jaime Martín Benito (Centro Nacional de Biotecnología, CNB-CSIC). Se optó por la clonación de la proteína completa en el vector pCR2.1-TOPO utilizando el kit *TOPO TA Cloning* (Invitrogen).

Para realizar la clonación del gen de gp17, se amplificó el gen utilizando los oligonucleótidos gp17.F.NcoI y gp17.R.HindIII. Para la subclonación, tras una restricción con las enzimas *NcoI* y *HindIII* del producto de PCR y del vector pETDuet durante 6 h, se procedió a la ligación durante 4 h a 20 °C.

Posteriormente se llevó a cabo una subclonación del inserto con las mismas enzimas de restricción en el vector pET30a(+), permitiendo expresar la proteína con una cola de seis histidinas en el extremo amino terminal. La secuencia de la cola es MHHHHHSSGLVPRGSGMKETAAAKFERQHMDSPDLGTDDD DKAM.

3.8.2 Clonación de los fragmentos carboxilo terminales de gp17

El diseño de los fragmentos se realizó basándose en las predicciones de los distintos dominios de la proteína realizadas por el Prof. A. Steven (Steven *et al.*, 1988). Se diseñaron cuatro fragmentos carboxilo terminales, empezando cada uno de ellos en los aminoácidos 281, 371, 431 y 461 respectivamente (Figura 3.5). Utilizando los correspondientes oligonucleótidos (gp17.F281.NcoI, gp17.F371.NcoI, gp17.F431.NcoI y gp17.F461.NcoI), se realizó la clonación en el vector pETDuet de cada uno de ellos, siguiendo el mismo protocolo que para el gen de la proteína completa. Posteriormente se llevó a cabo una subclonación del inserto

correspondiente al fragmento gp17(371-553) con las mismas enzimas de restricción en el vector pET30a(+), permitiendo expresar la proteína con una cola de seis histidinas en el extremo amino terminal, equivalente a la obtenida en el mismo vector para la proteína completa.

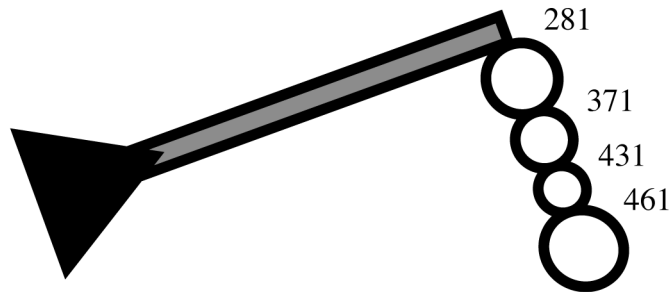


Figura 3.5: Esquema de la fibra de T7. Se señalan los aminoácidos en los que comienza cada uno de los nódulos, a partir de los cuales se han diseñado los diferentes fragmentos carboxilo terminales.

3.8.3 Expresión de gp17 sin cola de histidinas

El vector pETDuet-17 se utilizó para transformar la cepa de *E. coli* BL21(DE3). Se seleccionó una colonia aislada que se creció en 20 ml de cultivo líquido toda la noche a 37 °C. A la mañana siguiente se transfirió a un cultivo de 0,9 l con una dilución 1:100 y se dejó crecer a 37 °C hasta obtener una DO_{600} de entre 0,6 y 0,8. Se indujo la expresión de la proteína añadiendo IPTG hasta una concentración final de 1 mM y se dejó expresar a 16 °C durante 16 h. Los cultivos se centrifugaron 30 min a 3000 x g, se resuspendieron en 10 ml de tampón de lisis T7 (50 mM Tris-HCl pH 8,0, 4% (v/v) glicerol, 50 mM cloruro amónico, 2 mM EDTA, 150 mM cloruro sódico) y se congelaron hasta el momento de su purificación. Se analizó la expresión de la proteína por SDS-PAGE.

3.8.4 Purificación de gp17 sin cola de histidinas

Las bacterias se lisaron con tres ciclos de sonicación de 30 s espaciados por 30 s en hielo, a 70 kHz. La muestra se clarificó mediante una centrifugación de 30 min a 20000 x g a 4 °C. Al sobrenadante se le añadió un 10% (m/v) de sulfato amónico y se incubó 30 min a 4 °C. Posteriormente se centrifugó 30 min a 20000 x g a 4 °C. La proteína precipitada se resuspendió en tampón 10 mM fosfato sódico pH 6,8 y se dializó toda la noche frente al mismo tampón para eliminar los restos del tampón previo y del sulfato amónico precipitado.

Se realizó una cromatografía en columna de hidroxapatito con un gradiente de fosfato sódico entre 10 y 500 mM, a pH 6,8.

3.8.5 Expresión y purificación de gp17 con cola de histidinas

Se transformó la cepa de *E. coli* BL21(DE3) con el vector pET30-17 y se realizó la expresión del mismo modo que para la proteína sin cola de histidinas.

Las bacterias se lisaron pasándolas tres veces a través de un emulsificador C5 Avestin (Avestin Europe GmbH). Se centrifugó el lisado 30 min a 20000 x g a 4 °C y el sobrenadante se incubó 30 min en agitación con 0,5 ml de resina Ni-NTA (níquel-ácido nitriloacético) (Jena Bioscience) por cada litro de cultivo inicial. Se cargó en una columna y se realizó una elución con un gradiente de imidazol en tampón 50 mM Tris-HCl pH 8,5, 300 mM cloruro sódico, 50 mM imidazol (100, 150, 200, 250, 300, 400 y 1000 mM de imidazol). Se descartaron las fracciones de 50 y 100 mM y las fracciones restantes se dializaron frente a tampón A (10 mM Tris-HCl pH 8,5) durante toda la noche. A continuación se realizó una cromatografía de intercambio aniónico en una columna UNO-Q12 (Bio-Rad) utilizando tampón A y eluyendo con un gradiente lineal de cloruro sódico entre 0 y 1 M en el mismo tampón. Las fracciones que contenían la proteína se concentraron con concentradores Amicon Ultra 10 K (Millipore) y se realizaron tres lavados con tampón A para eliminar el cloruro sódico.

3.8.6 Expresión de los fragmentos carboxilo terminales de gp17 sin cola de histidinas

Los vectores pETDuet-17(281-553), pETDuet-17(371-553), pETDuet-17(431-553) y pETDuet-17(461-553) se utilizaron para transformar la cepa de *E. coli* BL21(DE3). Se procedió de la misma forma que con gp17 sin cola de histidinas, tal y como se describe en los apartados 3.8.3 y 3.8.4. Para el análisis de la expresión se lisaron las células con tres ciclos de sonicación de 30 s espaciados por 30 s en hielo a 70 kHz y se clarificó la muestra con una centrifugación de 30 min a 20000 x g a 4 °C. Se analizó la presencia de cada uno de los fragmentos tanto en el sobrenadante como en el pellet en geles del 15% y el 20%.

3.8.7 Purificación de los fragmentos carboxilo terminales gp17(371-553) y gp17(431-553) sin cola de histidinas

Las bacterias se lisaron pasándolas tres veces a través de un emulsificador C5 Avestin (Avestin Europe GmbH). Se centrifugó el lisado 40 min a 39000 x g a 4 °C y el sobrenadante se incubó 20 min con 0,2 ng de DNAsa, 0,2 ng de RNasa y 10mM MgSO₄. A continuación se centrifugó 20 min a 20000 x g a 4 °C y al sobrenadante se le añadió 2 M de sulfato amónico. Se incubó 30 min y se centrifugó 30 min a 39000 x g. El sobrenadante se filtró con un filtro de 22 µm y se cargó en una columna de fenil sefarosa (GE Healthcare).

Para el fragmento gp17(371-553), se concentraron con concentradores Amicon Ultra 10K las fracciones de la cromatografía anterior que contenían la proteína y se cargaron en una columna de intercambio aniónico UNO Q-12 utilizando tampón A y eluyendo con un gradiente lineal de cloruro sódico entre 0 y 1 M en el mismo tampón.

Para el fragmento gp17(431-553), se concentraron las fracciones de la cromatografía anterior que contenían la proteína y se cargaron en una columna de exclusión molecular S-100 (GE Healthcare) utilizando el tampón SE (10 mM Tris-HCl pH 8,5, 100 mM cloruro sódico, 1% (v/v) glicerol).

En ambos casos, para analizar la presencia de la proteína en las diferentes fracciones a lo largo de la purificación, se utilizó Tricina-SDS-PAGE en lugar de SDS-PAGE.

3.8.8 Expresión y purificación del fragmento gp17(371-553) con cola de histidinas

Se transformó la cepa de *E. coli* BL21(DE3) con el vector pET30-17(371-553) y se realizó la expresión del mismo modo que para el fragmento sin cola de histidinas. La lisis se realizó con tres ciclos de sonicación de 30 s espaciados por 30 s en hielo a 70 kHz. La purificación se realizó del mismo modo que para la proteína gp17 con cola de histidinas.

3.8.9 Tripsinización de gp17 y del fragmento gp17(371-553)

Las proteínas obtenidas de la expresión de los plásmidos pET30-17 y pET30-17(371-553) tras la purificación en la columna Ni-NTA se dializaron frente a tampón 20 mM bicarbonato amónico, 150 mM cloruro sódico y se incubaron a 37 °C durante

30 min. Se añadió a las muestras 7,5 µg de tripsina (Promega) por mg de proteína, incubándose hasta 120 min y recogiendo alícuotas a tiempos 0, 30 y 120 min. Los fragmentos obtenidos se identificaron por secuenciación amino terminal. Posteriormente se procedió a la purificación mediante cromatografía de intercambio iónico UNO-Q12, del mismo modo que para la proteína sin tripsinizar.

3.8.10 Microscopía electrónica de gp17

Las muestras que fueron utilizadas para microscopía electrónica se diluyeron en tampón 10 mM HEPES pH 7,5 y se dializaron frente a este mismo tampón toda la noche. Para estos experimentos se utilizó gp17 sin cola de histidinas, ya que permitiría ver la fibra tal y como se expresa en la célula, sin posibles artefactos por la presencia de la cola amino terminal. La concentración final de la muestra se ajustó a 0,5 mg/ml, realizando diluciones seriadas en el mismo tampón hasta que las partículas se visualizaron de forma independiente en la rejilla. Se utilizaron rejillas cubiertas con una capa de colodión y carbón que fueron sometidas a una descarga iónica e incubadas con la muestra. Se realizaron lavados con tampón para eliminar el exceso de muestra y a continuación se realizó la tinción con acetato de uranilo al 2% (m/v).

La preparación de las rejillas y toma de imágenes de microscopía electrónica fueron llevadas a cabo por el Dr. José Otero y el Dr. Guy Schoehn en el Institute de Biologie Structurale (IBS) de Grenoble (Francia).

3.8.11 Cristalización del fragmento gp17(371-553)

Se realizaron numerosos ensayos de cristalización utilizando la técnica de difusión de vapor en gota sentada, probando diferentes kits comerciales y a diferentes concentraciones de proteína, tanto para el fragmento completo como para el obtenido mediante tripsinización. Paralelamente, varias muestras de proteína fueron enviadas para realizar experimentos de cristalización de alto rendimiento (HTX) en el Laboratorio de Biología Molecular Europeo (EMBL) de Grenoble (Dimasi *et al.*, 2007). Dichos experimentos se realizaron a 20 °C con los kits *Crystal Screen*, *Crystal Screen 2*, *Crystal Screen Lite*, *PEG/Ion*, *MembFac*, *Natrix*, *Quick Screen*, *Grid Screens Ammonium Sulfate*, *Malonate*, *PEG 6000*, *PEG/LiCl* y *MPD* (Hampton Research) y los kits *JCSG+* y *PACT* (Qiagen).

Las condiciones en las que se observaron cristales se reprodujeron en el laboratorio con un tamaño de gota mayor, utilizando la misma técnica de difusión de

vapor en gota sentada. Para ello se utilizaron las placas de cristalización *CrysChem* (Hampton Research) con 0,5 ml en el reservorio y gotas de 2 μ l de proteína mezclados con el mismo volumen de solución del reservorio.

Los cristales obtenidos en gotas cuya condición de cristalización era 0,1 M Tris-HCl pH 8,5, 0,2 M trimetilamina-N-óxido y 20-25% (m/v) PEG MME (polietilenglicol mono metil éter) 2000 fueron congelados en nitrógeno líquido y utilizados para la recogida de datos.

Para determinar las fases, los cristales fueron impregnados con compuestos de átomos pesados. Estos compuestos fueron añadidos en polvo al reservorio de varios pocillos de cristalización para generar una solución saturada que se dejó equilibrar toda la noche. Al día siguiente, 1 μ l de esta solución saturada fue añadido a la gota y los cristales fueron congelados tras diferentes tiempos de incubación. Se recogieron cristales tras varios minutos, varias horas y un día.

Los datos fueron recogidos en la línea BM14 del *European Synchrotron Radiation Facility* (ESRF, Grenoble, Francia) con un detector MAR Research CCD 225 (MAR Research).

Se recogieron sets de datos a diferentes longitudes de onda para un derivado con mercurio (pico, punto de inflexión, remoto de alta energía y remoto de baja energía).

Los diferentes sets de datos fueron integrados con MOSFLM (Battye *et al.*, 2011) y escalados con SCALA (Evans, 2011). A continuación se utilizó el programa AUTOSHARP (Vonrhein *et al.*, 2007), que identificó tres mercurios con los programas SHELXC/D (Sheldrick, 2008). Las fases fueron refinadas usando los datos entre 20 y 2,7 Å con AUTOSHARP, que añadieron y eliminaron átomos de mercurio hasta obtener finalmente nueve. El aplanamiento de solvente se realizó con SOLOMON (Abrahams y Leslie, 1996) y la construcción automatizada del modelo con BUCCANEER (Cowtan, 2006), a 2,7 Å de resolución. Este primer modelo se utilizó para un reemplazamiento molecular a una resolución mayor con un set de datos nativo, utilizando el programa MOLREP (Vagin y Isupov, 2001). El modelo se introdujo en ARP-WARP (Langer *et al.*, 2008) para realizar el autotrazado a una resolución de 1,9 Å. El ajuste del modelo, adición de aminoácidos extra y de moléculas de solvente se realizó con COOT (Emsley *et al.*, 2010), y el refinamiento con REFMAC (Murshudov *et al.*, 2011).

Paralelamente, un modelo parcialmente refinado se utilizó para resolver la estructura de un cristal obtenido en 0,2 M citrato amónico y 20% (m/v) PEG 3350

mediante reemplazamiento molecular con el programa MOLREP y siguiendo el mismo procedimiento que con la otra forma cristalina.

La validación del modelo se realizó con el programa MOLPROBITY (Chen *et al.*, 2010). Para el análisis de la superficie expuesta y para el cálculo de la energía de disociación se utilizó el servidor PISA (Krissinel y Henrick, 2007). Ambas estructuras han sido depositadas en el PDB con los códigos 4a0t (la forma cristalina $P2_12_12_1$) y 4a0u (la forma cristalina $C222_1$).

La comparación de la estructura obtenida con otras estructuras se ha realizado con el servidor DALI (Holm y Rosenstrom, 2010), y las imágenes se han realizado con PyMOL (Schrödinger, LLC).

3.9 Ensayos de inhibición del fago T7

Para realizar los ensayos de inhibición de la infección por T7 se crecieron células de la cepa de *E. coli* BL21 (excepto cuando se especifique otra cepa) en LB durante toda la noche. A la mañana siguiente se diluyó el cultivo 5:100 en LB hasta alcanzar una DO_{600} de 0,4. Las células se mantuvieron en hielo hasta su uso. Para las pruebas de competición, se incubaron 8×10^7 células con 1,5 mg de proteína o con la cantidad indicada de LPS (lipopolisacáridos) durante 1 hora a 4 °C. A continuación, se añadió la reacción a una placa Petri vacía, añadiendo también una gota con 100-150 pfu (unidades formadoras de placa, virus), manteniendo ambas gotas independientes y añadiendo sobre ambas de forma inmediata 10 ml de LB-topagar. Se homogenizó el contenido de la placa, se dejó solidificar y se incubó a 37 °C toda la noche. A la mañana siguiente se contaron las placas de lisis en cada placa Petri.

Se realizaron diferentes pruebas:

- **Ensayo 1:** Se realizaron pruebas con diferentes diluciones del stock de fagos para calcular cual era la dilución correcta para utilizar en los siguientes ensayos. Se probaron las diluciones 10^{-9} , 10^{-10} , 10^{-11} en las cepas BL21 y BL21(DE3). Se realizaron controles en los que se sembraron bacterias sin añadir fagos, tampón TMS sin bacterias ni fagos y bacterias con TMS con 10^{-10} fagos para ambas cepas.
- **Ensayo 2:** Se realizó con la cepa BL21 y se contaron las placas de lisis obtenidas tras haber incubado las células con BSA y con gp17(371-553). Los fagos se utilizaron a una dilución 10^{-10} . Para cada una de las reacciones se repitió el experimento tres veces y se calculó la media.

- **Ensayo 3:** Para confirmar los resultados obtenidos previamente, se repitió el ensayo 2 con BSA y gp17(371-553), preparando dos reacciones de cada proteína.
- **Ensayo 4:** Se probaron dos muestras diferentes de gp17(371-553) por duplicado y una muestra de gp17(371-553) tripsinizada.
- **Ensayo 5:** En lugar de utilizar proteína purificada se utilizaron lisados bacterianos de la cepa BL21(DE3) transformada con los plásmidos para gp17(371-553) con cola de histidinas y sin ella. Se comprobó la expresión de ambas y se determinó que la proteína con cola de histidinas se expresaba seis veces más que la otra, por lo que se ajustó la cantidad de muestra añadida a este factor. Se hicieron pruebas con 600, 300 y 150 µl de lisado en el caso de la proteína sin cola de histidinas y 100, 50 y 25 µl en el caso de la proteína con cola de histidinas, todas por duplicado.
- **Ensayo 6:** Se realizaron reacciones incubando con los LPS O55 y O111. En ambos casos se realizó una reacción con 50 µg y otra con 5 µg. Se midió el diámetro de todas las placas de lisis obtenidas, así como de las obtenidas en el control sin LPS.
- **Ensayo 7:** Se repitió el ensayo anterior añadiendo en este caso 100 µg del LPS correspondiente y del control, ambos ensayos por triplicado.
- **Ensayo 8:** Se probó otro LPS, el EH100, realizando dos repeticiones. Se hicieron controles con los dos LPS utilizados en los ensayos anteriores. En todas las pruebas se añadieron 200 µg del LPS correspondiente. Se incluyeron reacciones con gp17(371-553) con y sin cola de histidinas por duplicado.
- **Ensayo 9:** Se realizaron pruebas con 200, 100, 10, 1, 0,1 0,01 y 0,001 µg del LPS EH100.
- **Ensayo 10:** Se realizaron pruebas con 5, 2,5, 1,25, 0,75 y 0,375 µg del LPS EH100.

En todos los ensayos se realizaron los controles descritos en el ensayo 1.

3.10 Purificación de colas del bacteriófago T5

3.10.1 Crecimiento y purificación de fagos (mutante *T5amD20*)

El mutante *T5amD20* fue proporcionado por la Dra. Pascale Boulanger (IBBMC-CNRS). El proceso de obtención de colas del bacteriófago se llevó a cabo en su laboratorio en el Instituto de Bioquímica y Biofísica Molecular y Celular (IBBMC-CNRS, Orsay, Francia) durante una estancia breve.

Para crecer los fagos se emplea una cepa permisiva (*E. coli* CR63) que permite la correcta formación de la partícula viral. Los mutantes *amber* se caracterizan por la mutación de un codón CAG por un codón UAG. Cuando la proteína mutada se expresa en una cepa no supresora (también llamada no permisiva), el codón UAG se interpreta como un codón de parada y la proteína deja de traducirse. Si por el contrario la expresión se realiza en una cepa supresora (o permisiva), en el lugar correspondiente se inserta un aminoácido, que varía dependiendo del tRNA que tenga la cepa supresora. En el caso del mutante usado en este trabajo, *T5amD20*, la mutación se encuentra en el gen que codifica la proteína mayoritaria de la cápside (pb8), por lo que al multiplicarlo en una cepa no permisiva, la cápside no puede ensamblarse y se obtienen sólo las colas con sus correspondientes fibras.

Se realizó un cultivo en LB de la cepa CR63 durante toda la noche a partir de una colonia aislada y la mañana siguiente se prepararon cultivos en LB suplementado con 1 mM de cloruro cálcico y 1 mM de cloruro magnésico, añadiendo las células en una dilución 1:100. Se dejaron crecer hasta una DO_{600} de 0,3 y en ese momento se añadieron fagos *T5amD20* al cultivo a una multiplicidad de infección (MOI) de 10^{-2} y se dejaron lisis hasta alcanzar una DO_{600} de 0,1. Entonces se añadió DNasa (10 µg/ml) y 0,2 ml de cloroformo por litro de cultivo. Se incubó 30 min a 37 °C. A continuación, se añadió 500 mM de cloruro sódico hasta su completa disolución y se centrifugaron los cultivos 15 min a 8000 x g. Al sobrenadante se le añadió un 8% (m/v) de PEG 6000 y se incubó toda la noche a 4 °C.

A la mañana siguiente se centrifugaron los cultivos a 8000 x g durante 20 min y se resuspendió el pellet en tampón P (10 mM Tris-HCl pH 7,4, 150 mM cloruro sódico, 1 mM cloruro cálcico, 1 mM cloruro magnésico). Se centrifugó de nuevo a 6000 x g durante 20 min para eliminar los restos celulares.

3.10.2 Purificación de fagos mediante gradiente de cloruro de cesio

Se prepararon gradientes discontinuos para la ultracentrifugación, con tres capas de cloruro césico a diferentes densidades: 1,6, 1,5 y 1,4 (g/ml); añadiendo sobre ellas la capa de fagos purificados previamente. Cada una de las cuatro capas tenía un volumen de 2 ml. Se centrifugó durante 2 h a 68400 x *g*. Al extraer los tubos se observó una capa de color lechoso a una altura aproximada de dos tercios del volumen total del gradiente que se corresponde con los fagos. Se eliminaron las capas superiores con una pipeta Pasteur y se recogió la capa de fagos. Los fagos purificados se dializaron toda la noche frente a tampón P.

3.10.3 Crecimiento y purificación de colas

Para extraer las colas del fago T5, se utilizó la cepa no permisiva *E. coli* F. Se utilizó el mismo protocolo que para el crecimiento en la cepa permisiva CR63, pero en este caso los fagos se añadieron a una MOI de 8. Una vez lisadas las células, se llevó a cabo el mismo protocolo de purificación, exceptuando el gradiente de cloruro de cesio, que se sustituyó por un gradiente de glicerol.

Para la purificación por gradiente de glicerol, se preparó un gradiente discontinuo de glicerol con soluciones de 2 ml al 40% (v/v), 30%, 20% y 10%, preparados en tampón G (20 mM Tris-HCl pH 8, 200 mM cloruro sódico), sobre los que se depositó la capa de colas. Se centrifugó 2 h a 68400 x *g*. El contenido del tubo se separó en fracciones de 1 ml que se analizaron por SDS-PAGE.

Los tubos en los que estaban presentes las colas se diluyeron en tampón C (20 mM Tris-HCl pH 7,5, 1 mM cloruro cálcico, 1 mM cloruro magnésico) y se realizó un último paso de purificación por cromatografía de intercambio iónico en una columna HiTrap (GE Healthcare) con el mismo tampón C y realizando un gradiente de cloruro sódico entre 75 y 1000 mM.

3.10.4 Microscopía electrónica y procesamiento digital de las colas de T5

Las fracciones de las colas de T5 recogidas durante la purificación se analizaron mediante microscopía electrónica de transmisión. Para ello, entre 2 y 5 µl de muestra fueron adsorbidos en una rejilla electrónica con carbón-colodión previamente ionizada mediante descarga iónica (K100X, Emitech), lavada dos veces con agua y teñidas con acetato de uranilo al 2% (m/v). Las muestras fueron visualizadas en un microscopio Tecnai G² FEG 200 (FEI) estabilizado a 200 kV y las imágenes se

registraron en una cámara CCD (*Charged Coupled Device*) FEI Eagle 4k a una magnificación nominal 67,873X sobre el detector (2,21 Å/píxel).

Las operaciones de procesamiento general de imágenes fueron llevadas a cabo con los paquetes de procesamiento *Xmipp* (Marabini *et al.*, 1996) y Bsoft (Heymann, 2001; Heymann y Belnap, 2007). Mediante el programa *helixboxer* del paquete EMAN2 (Tang *et al.*, 2007) fueron seleccionadas manualmente 1129 imágenes, seleccionando únicamente la parte distal de la fibra, que fueron posteriormente extraídas y normalizadas con *Xmipp*. El rango de desenfoque fue determinado con el programa *ctffind3* (Mindell y Grigorieff, 2003) mediante el cálculo de la media radial de la transformada de Fourier promedio de la micrografía, lo que fue empleado para invertir las fases de cada micrografía entre las frecuencias adecuadas de la función de transferencia de contraste. Las imágenes fueron alineadas y clasificadas bidimensionalmente mediante el método de máxima verosimilitud (Scheres *et al.*, 2005). El modelo 3D fue generado aplicando simetría rotacional sobre el eje longitudinal de la fibra.

La reconstrucción 3D fue subdividida en siete “dominios” basados en los estrechamientos a lo largo del volumen de la fibra. Teniendo en cuenta la predicción de estructura secundaria de la proteína, se puede estimar que los 206 primeros aminoácidos forman la parte proximal de la fibra, por lo que se ha dividido el resto de la secuencia, asumiendo una distribución lineal. Se ha calculado el volumen total del modelo obtenido y se ha calculado el volumen de cada uno de los dominios, lo que ha permitido extrapolar que región de la secuencia estaría comprendida en este dominio. Posteriormente, el ajuste de la estructura en el interior del volumen de microscopía electrónica se llevó a cabo utilizando el programa *collage* de la suite *Situs* (Wriggers, 2010).

La obtención de imágenes y su procesado fue llevada a cabo en colaboración con el Dr. José Ruiz Castón y el Dr. Daniel Luque, del Departamento de Estructura de Macromoléculas del Centro Nacional de Biotecnología (CNB-CSIC).

3.11 La proteína pbl

3.11.1 Predicción de estructura secundaria de la fibra de T5

La predicción de estructura secundaria de la fibra de T5 se ha realizado con el servidor PsiPred v3.3 (Jones, 1999; Buchan *et al.*, 2010). Este tipo de predicciones

son sólo orientativas, pero ya que en este caso no se tiene ninguna información sobre la posible estructura de la fibra, permite predecir las regiones desordenadas, y decidir que oligonucleótidos se diseñan para clonar diferentes fragmentos carboxilo terminales.

3.11.2 Clonación de pb1 y de los fragmentos carboxilo terminales

Para realizar la clonación de las diferentes construcciones de pb1 se amplificó la secuencia nucleotídica correspondiente a cada fragmento con los oligonucleótidos que se describen en la Tabla 3.1. Para los tres primeros fragmentos se realizó la restricción con las enzimas *Bam*HI y *Not*I y para los restantes con *Bam*HI y *Hind*III, todas ellas durante 30 min empleando enzimas FastDigest (Thermo Fisher). Del mismo modo, se cortaron los vectores pET28c(+) (para los siete primeros, diseñados exclusivamente en base a la predicción de estructura secundaria) y pET28a(+) (para los tres últimos, diseñados teniendo en cuenta los resultados obtenidos por microscopía electrónica). Los productos se ligaron durante 3 h a 20 °C.

Tabla 3.1: Oligonucleótidos necesarios para amplificar cada uno de los fragmentos diseñados de pb1.

Fragmento	Oligonucleótido 1	Oligonucleótido 2
1-1396	pb1.F1.BamHI	pb1.R.NotI
224-1396	pb1.F224.BamHI	pb1.R.NotI
413-1396	pb1.F413.BamHI	pb1.R.NotI
674-1396	pb1.F674.BamHI	pb1.R.HindIII
693-1396	pb1.F693.BamHI	pb1.R.HindIII
794-1396	pb1.F794.BamHI	pb1.R.HindIII
1106-1396	pb1.F1106.BamHI	pb1.R.HindIII
728-1396	pb1.F728.BamHI	pb1.R.HindIII
835-1396	pb1.F835.BamHI	pb1.R.HindIII
970-1396	pb1.F970.BamHI	pb1.R.HindIII

Para los fragmentos pb1(1-1396), pb1(224-1396) y pb1(413-1396) no se obtuvieron clones positivos. Con el resto, se procedió a analizar su expresión. Los vectores obtenidos se utilizaron para transformar la cepa de *E. coli* BL21(DE3). Se seleccionó una colonia aislada que se creció en 20 ml de cultivo líquido toda la noche a 37 °C. A la mañana siguiente se transfirió a un nuevo cultivo de 20 ml con una dilución 1:100 y se dejó crecer a 37 °C hasta obtener una DO₆₀₀ de 0,6-0,8. Se indujo la expresión añadiendo 1 mM de IPTG y se dejó expresar a 37 °C durante 3 h.

Los cultivos se centrifugaron a 3000 x g durante 10 min y se resuspendieron en tampón de lisis T5 (50 mM Tris-HCl pH 8, 500 mM cloruro sódico, 10% (v/v) glicerol).

3.11.3 Purificación de los fragmentos pb1(794-1264), pb1(1106-1264) y pb1(970-1264).

Las bacterias se lisaron en hielo con tres ciclos de sonicación de 30 s espaciados por 30 s en hielo a 70 kHz. El lisado se centrifugó 30 min a 20000 x g a 4 °C y el sobrenadante se incubó 30 min en agitación suave con 0,5 ml de resina Ni-NTA por cada litro de cultivo inicial. Se cargó en una columna y se realizó una elución con un gradiente de imidazol en tampón 50 mM Tris-HCl pH 8,5, 300 mM cloruro sódico, 50 mM imidazol (100, 150, 200, 250, 300, 400 y 500 mM de imidazol). Se recogieron las fracciones de entre 150 y 400 mM de imidazol y se dializaron frente a tampón A (10 mM Tris-HCl pH 8,5). A continuación se realizó una cromatografía de intercambio aniónico en una columna Resource Q-6 (GE Healthcare) utilizando tampón A y eluyendo con un gradiente lineal de cloruro sódico entre 0 y 1 M en el mismo tampón. Las fracciones que contenían la proteína se concentraron con concentradores Amicon Ultra 10K y se realizaron tres lavados con tampón A para eliminar el cloruro sódico. Se utilizó el mismo proceso para las construcciones pb1(1106-1264) y pb1(970-1264).

3.11.4 Expresión y purificación del fragmento pb1(970-1264) con selenio-metionina

Puesto que no se dispone de la estructura de una proteína con secuencia homóloga resuelta, en caso de obtener cristales de esta proteína no se podría llevar a cabo un reemplazamiento molecular, y sería necesario introducir átomos pesados. Teniendo en cuenta este punto, se decidió llevar a cabo la expresión de este fragmento con selenio en lugar de azufre en las metioninas. Se transformó la cepa B834(DE3) con el plásmido que contenía el inserto pb1(970-1396). Se seleccionaron colonias obtenidas en placas de LB-agar y se crecieron toda la noche en 100 ml de medio para selenio-metionina (Molecular Dimensions) suplementado con metionina. A la mañana siguiente se transfirió a un litro de cultivo con el mismo medio, suplementado en este caso con selenio-metionina. Se dejó crecer hasta alcanzar una DO_{600} de 0,4, se indujo a 1 mM de IPTG y se dejó expresar 3 h a 37 °C. La

purificación de este fragmento expresado con selenio-metionina se llevó a cabo con el mismo protocolo que en el caso de la expresión en BL21(DE3).

3.11.5 Cristalización de los fragmentos de pb1

Se llevaron a cabo ensayos de cristalización con todas las muestras purificadas que se obtuvieron de cada uno de los fragmentos. Se obtuvieron cristales del fragmento pb1(794-1264) en una gota cuya condición de cristalización contenía MES pH 6,5 y sulfato amónico, por lo que se replicaron estos ensayos de cristalización aumentando el tamaño de las gotas. Para hacer un muestreo de condiciones similares a aquella en la que se habían obtenido cristales, se probaron otros tampones (Tris, HEPES) a pH entre 6 y 8,5, y también diferentes concentraciones de precipitante, variando la concentración de sulfato amónico entre 0,2 y 2,5 M. Por último, se diseñaron condiciones en las que se intercambió el sulfato amónico total o parcialmente por otros precipitantes (PEG 4000, PEG 8000 y ácido málico) y se añadieron diferentes crioprotectores a las condiciones (glicerol, PEG 400).

Las diferentes condiciones diseñadas en función de los resultados obtenidos con el fragmento pb1(794-1264) fueron probadas también con el fragmento pb1(970-1264), tanto en su forma nativa como en el derivado con selenio-metionina.

Los datos del fragmento pb1(970-1264) con selenio-metionina fueron recogidos en la línea I02 del sincrotrón *Diamond* (Oxford), con un detector Pilatus3 6M (Dectris), recogiendo los datos a la longitud de onda correspondiente al pico del selenio. Los datos fueron integrados y escalados utilizando el paquete *xia2* (Evans, 2006; Kabsch, 2010; Winter, 2010). A continuación se utilizó el programa AUTOSHARP (Vonrhein *et al.*, 2007), que identificó 15 selenios utilizando los programas SHELXC/D (Sheldrick, 2008). Las fases fueron refinadas con AUTOSHARP y el aplanamiento de solvente se realizó con SOLOMON (Abrahams y Leslie, 1996). Posteriormente se llevó a cabo el autotrazado del modelo con PHENIX (Adams *et al.*, 2010). El ajuste del modelo, adición de aminoácidos extra y de moléculas de solvente se realizó con COOT (Emsley *et al.*, 2010) y el refinamiento con REFMAC (Murshudov *et al.*, 2011). La validación del modelo se realizó con el programa MOLPROBITY (Chen *et al.*, 2010). Para el análisis de la superficie expuesta y el cálculo de la energía de disociación se utilizó el servidor PISA (Krissinel y Henrick, 2007).

Teniendo en cuenta que el extremo carboxilo terminal de la proteína se proteoliza liberándose el dominio chaperona y que este dominio tiene una alta homología de secuencia con la chaperona de la espina del fago K1F, se realizó un *docking* de la estructura obtenida con la estructura de dicha chaperona, para lo que se utilizó el servidor ZDOCK (Pierce *et al.*, 2011). La comparación de la estructura obtenida con otras estructuras se ha realizado con el servidor DALI (Holm y Rosenstrom, 2010), y las imágenes se han realizado con PyMOL (Schrödinger, LLC).

3.12 Alineamiento de secuencias

3.12.1 La fibra de T7

Se ha realizado una búsqueda exhaustiva de secuencias homólogas a gp17 con el objetivo de localizar otras proteínas que pudieran descender de un antecesor común. Así, una comparación de gp17 con proteínas homólogas, en combinación con la estructura conocida, puede desvelar qué aminoácidos podrían desarrollar una función importante, por ejemplo, en la unión al receptor.

Como se pretendían encontrar secuencias que fuesen similares en la mayor parte de su secuencia pero que mostrasen diferencias en otras regiones, así como enfatizar la posible presencia de deleciones o de dominios conservados, se decidió utilizar la técnica de alineamiento local.

La búsqueda se realizó usando la herramienta BLAST (Altschul *et al.*, 1990) utilizando el algoritmo *blastp* que compara la secuencia de entrada con la base de datos de secuencias de proteínas no redundantes. Para ello se lleva a cabo un fragmentado de la secuencia en palabras de tres aminoácidos, que son los que se buscan en la base de datos, y los resultados se muestran en función de una puntuación asignada utilizando la matriz BLOSUM62.

De los resultados obtenidos, se utilizaron para el alineamiento aquellos cuya identidad máxima con gp17 era superior al 75%. A continuación se realizó un alineamiento con el servidor de Clustal Omega (Sievers *et al.*, 2011). Para la representación gráfica se ha utilizado UCSF Chimera (Pettersen *et al.*, 2004).

3.12.2 La fibra de T5

En el caso de pb1 del fago T5, el proceso seguido fue similar al realizado con gp17. La principal diferencia entre ambos radica en el criterio de selección, ya que al no existir un salto tan claro en identidad de secuencia como en el caso de gp17, se

Estructura de fibras de bacteriófagos

seleccionaron aquellas secuencias cuya identidad era superior al 50%, para las que el valor de E era 0, teniendo en cuenta por lo tanto aquellas cuya puntuación máxima era 666 o superior. El resto del proceso realizado ha sido equivalente al de gp17.

4 Resultados y discusión

4.1 La proteína gp17

El bacteriófago T7 tiene seis fibras, que son homotrímeros paralelos y elongados de gp17 y que se unen a la cola del fago por su extremo amino terminal. Estas fibras son responsables del primer reconocimiento de la célula hospedadora por parte del fago.

4.1.1 Clonaciones

Con el fin de obtener información estructural sobre gp17, se clonaron fragmentos de DNA codificando gp17 y sus fragmentos carboxilo terminales tal y como se describe en Materiales y Métodos, obteniendo los vectores pETDuet-gp17, pET30-gp17, pETDuet-gp17(281-553), pETDuet-gp17(371-553), pETDuet-gp17(431-553), pETDuet-gp17(461-553) y pET30-gp17(371-553) (Figura 4.1). En los casos en los que se utilizó el vector pET30a(+), los vectores codifican además una cola de histidinas en el extremo amino terminal de la secuencia clonada para facilitar la purificación de las proteínas recombinantes.

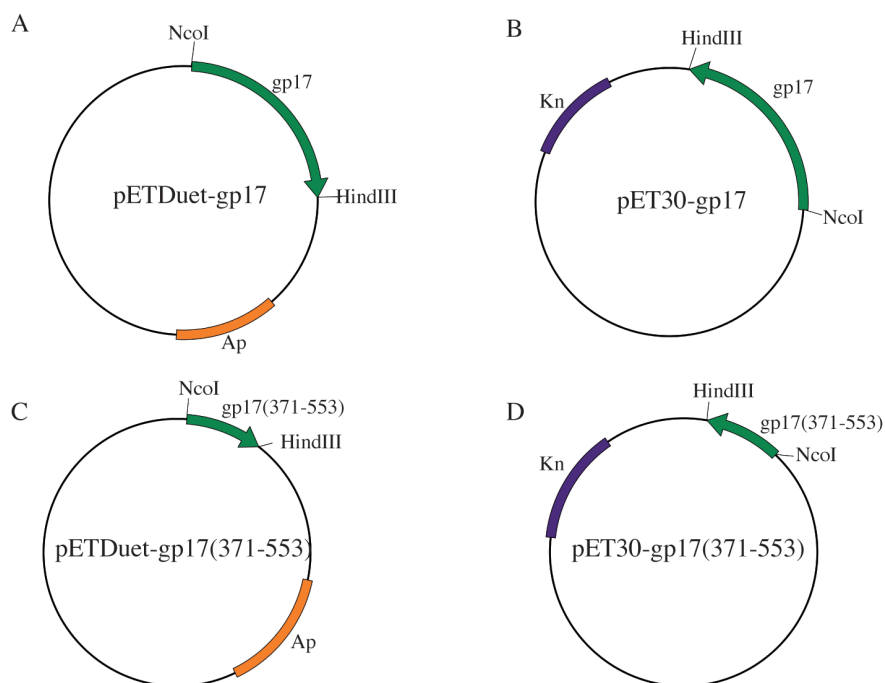


Figura 4.1: Modelos de vectores construidos para la expresión de gp17 o sus fragmentos. (A) pETDuet-gp17, (B) pET30-gp17, (C) pETDuet-gp17(371-553) y (D) pET30-gp17(371-553). Se ha seleccionado a modo representativo sólo uno de los cuatro fragmentos carboxilo terminales clonados, eligiendo aquel con el que más resultados se han obtenido.

Los fragmentos carboxilo terminales de gp17 a expresar, han sido elegidos teniendo en cuenta fundamentalmente los resultados obtenidos por Steven (Steven *et al.*, 1988). En este trabajo, mediante un promediado de microscopía electrónica, se predijeron los dominios de gp17. Según su predicción, la fibra presenta dos puntos de flexibilidad: uno en la unión a la cola del fago y otro entre las partes proximal y distal, la "rodilla". La primera parte sería la zona de unión a la cola del fago y abarcaría hasta el aminoácido 150. A continuación, se encontraría el asta o parte proximal, que terminaría en el aminoácido 267 o 280, y parte distal de la fibra que estaría formada por cuatro nódulos. Basándose en la frecuencia de los aminoácidos se predijo un modelo, teniendo en cuenta que las secuencias entre los residuos 275-375 y 460-553 eran relativamente hidrofóbicas, por lo que podrían formar estas regiones nodulares, y que existían pequeñas secuencias que favorecían la formación de hélices α , entre los aminoácidos 366-377, 421-433, 457-469, que podrían formar hélices superenrolladas uniendo los diferentes nódulos. Tomando como base este trabajo, cada uno de los fragmentos clonados comienza aproximadamente en estas regiones, de forma que el fragmento gp17(281-553) tiene los cuatro dominios, el gp17(371-553) tiene tres, el gp17(431-553) tiene dos y el gp17(461-553) tiene sólo el más distal. Además, se eligieron regiones para el comienzo de cada fragmento que estuvieran situados en zonas predichas como más desordenadas, para intentar que no afectase al plegamiento de la proteína resultante.

4.1.2 Expresión y purificación de los fragmentos de gp17

Todos los experimentos de expresión se llevaron a cabo a 16 °C para facilitar el plegamiento de gp17. Los primeros intentos de purificación se realizaron con la proteína obtenida del vector pETDuet-gp17. La purificación consistió en dos pasos: una precipitación con sulfato amónico y una cromatografía en columna de hidroxiapatito. En el primer paso ya se observó que la proteína precipitaba con un 10% (m/v) de sulfato amónico, lo que indicaba que podría tener tendencia a la agregación. Tras el segundo paso de purificación, se observó que todavía existían contaminantes en la muestra y se había obtenido muy poca proteína (menos de 1 mg por litro de cultivo), pero se decidió llevar a cabo los primeros ensayos de cristalización. Al concentrar la proteína purificada se confirmó que efectivamente tendía a agregar, lo que se puede explicar teniendo en cuenta la morfología de la proteína. Al ser una proteína fibrilar con un punto de flexibilidad en la mitad, la

"rodilla", es razonable que se produzcan agregados con facilidad si la proteína no está suficientemente pura y en un entorno adecuado, y también explica que en la precipitación con sulfato amónico el porcentaje necesario fuese bajo.

Como alternativa, intentando obtener una cantidad de proteína mayor que permitiese realizar más ensayos, se optó por la purificación de la proteína expresada por el vector pET30-gp17. Al poseer la cola de histidinas la purificación es más rápida y se obtiene proteína más pura. Así pues, tras una primera cromatografía de afinidad se obtuvo proteína mucho más pura (Figura 4.2A), a la que se le eliminaron los restos de contaminantes mediante una segunda cromatografía, en este caso de intercambio iónico (Figura 4.2B). En el gel se puede observar la presencia una banda de alto peso molecular cuando no se hierva la muestra, lo que confirma la presencia de la proteína trimérica resistente al SDS (Figura 4.2C). Los resultados obtenidos en este caso son mejores, pudiendo purificar entre 3 y 6 mg de proteína por litro de cultivo. De nuevo se llevaron a cabo ensayos de cristalización, pero no se obtuvieron cristales. Ya que una de las razones por las que es posible que la proteína no cristalizase era la flexibilidad de la rodilla, se procedió a realizar los ensayos con los fragmentos carboxilo terminales.

Una vez obtenidos los vectores de expresión para los fragmentos carboxilo terminales basados en el vector pETDuet, se comprobó el grado de expresión de cada uno de ellos. Al analizar en un gel SDS-PAGE la expresión obtenida con los vectores pETDuet-gp17(281-553) y pETDuet-gp17(461-553) no se obtuvieron bandas correspondientes a la sobreexpresión, pero sí se obtuvieron para los fragmentos expresados de los vectores pETDuet-gp17(371-553) y pETDuet-gp17(431-553). Con estos dos últimos, se procedió a la purificación. En ambos casos se eliminaron parte de las proteínas celulares mediante precipitación con sulfato amónico y se realizó una cromatografía de interacción hidrofóbica, pero el grado de pureza en ambos casos no era suficiente, por lo que se optó por realizar un paso extra en la purificación, diferente para cada fragmento. El fragmento gp17(371-553) se purificó mediante una cromatografía de intercambio aniónico, y el fragmento gp17(431-553) mediante una cromatografía de exclusión molecular. Se utilizaron técnicas diferentes para los dos fragmentos en función de cual de ellas permitía obtener un grado de pureza mayor. En ambos casos se consiguió obtener pequeñas cantidades de proteína pura para llevar a cabo ensayos de cristalización, pero las muestras tenían tendencia a

Estructura de fibras de bacteriófagos

agregar, dando lugar a precipitados cuando se intentaba incrementar la concentración de la proteína, por lo que no se obtuvieron cristales.

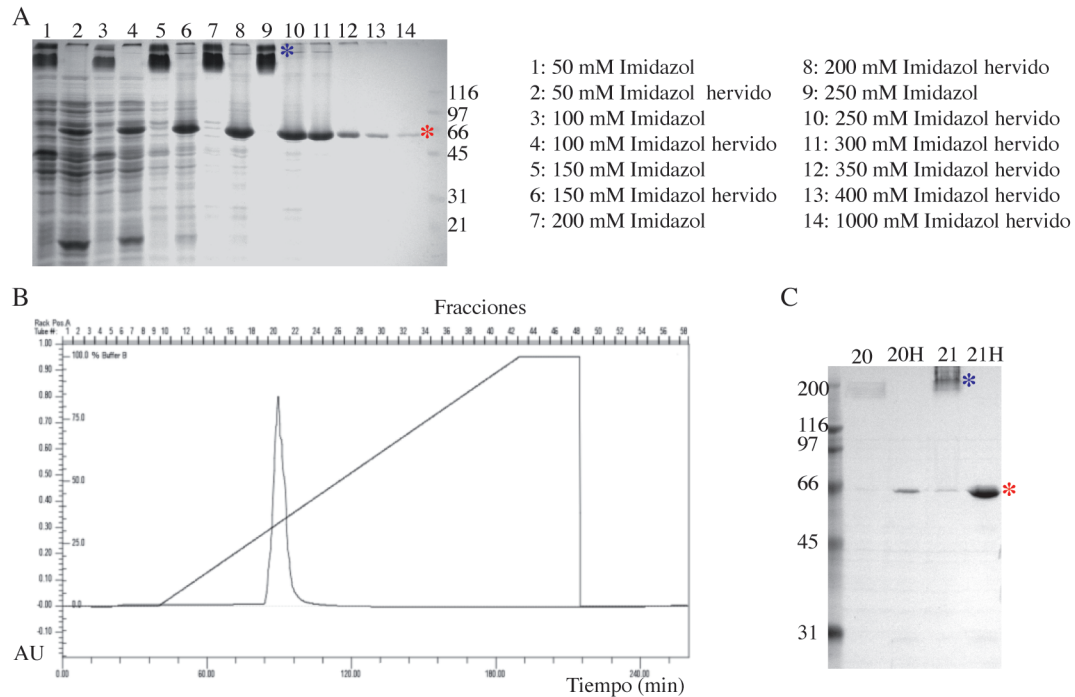


Figura 4.2: Purificación de gp17 con cola de histidinas. (A) Gel obtenido tras la purificación en columna de afinidad de gp17. Se muestra que la proteína comienza a eluir a partir de 150 mM de imidazol, y la forma trimérica al no hervir las muestras. (B) Cromatograma de intercambio aniónico. El pico eluye a 340 mM de cloruro sódico. (C) Gel de dos fracciones tomadas en la elución de la columna de intercambio iónico. En las muestras no hervidas se puede observar el trímero resistente al SDS (asterisco azul) y en las hervidas (H) el monómero (asterisco rojo).

Del mismo modo que con gp17 completa, se decidió realizar la purificación de un fragmento con cola de histidinas que había sido clonado en el vector pET30a(+). Se eligió el fragmento gp17(371-553) ya que de los cuatro fragmentos diseñados era el que mostraba mejor grado de expresión, y en las purificaciones previas era también con el que se obtenían mejores resultados.

La purificación del fragmento obtenido de la expresión del vector pET30-gp17(371-553) se realizó mediante una cromatografía de afinidad en la que la proteína eluye concentraciones entre 150 y 400 mM de imidazol (Figura 4.3A). Al igual que gp17 completa, se observó que este fragmento es resistente a SDS ya que si no se hierven las muestras antes de la electroforesis, gran parte migra en forma

trimérica. En el gel también se puede observar cómo el cambio de vector de expresión provoca un gran aumento en la cantidad de proteína soluble obtenida, ya que la cola de histidinas favorece la estabilidad de la proteína que, cuando estaba expresada sin cola de histidinas, podría estar parcialmente desplegada en su extremo amino terminal, lo que favorecería la agregación que se estaba observando, y que impedía incrementar la concentración de proteína en la muestra.

Para mejorar el grado de pureza antes de la cristalización se realizó una cromatografía de intercambio aniónico. El fragmento gp17(371-553) eluye a una concentración de 170 mM de cloruro sódico en un único pico (Figura 4.3B y C). La proteína obtenida se concentró para realizar ensayos de cristalización.

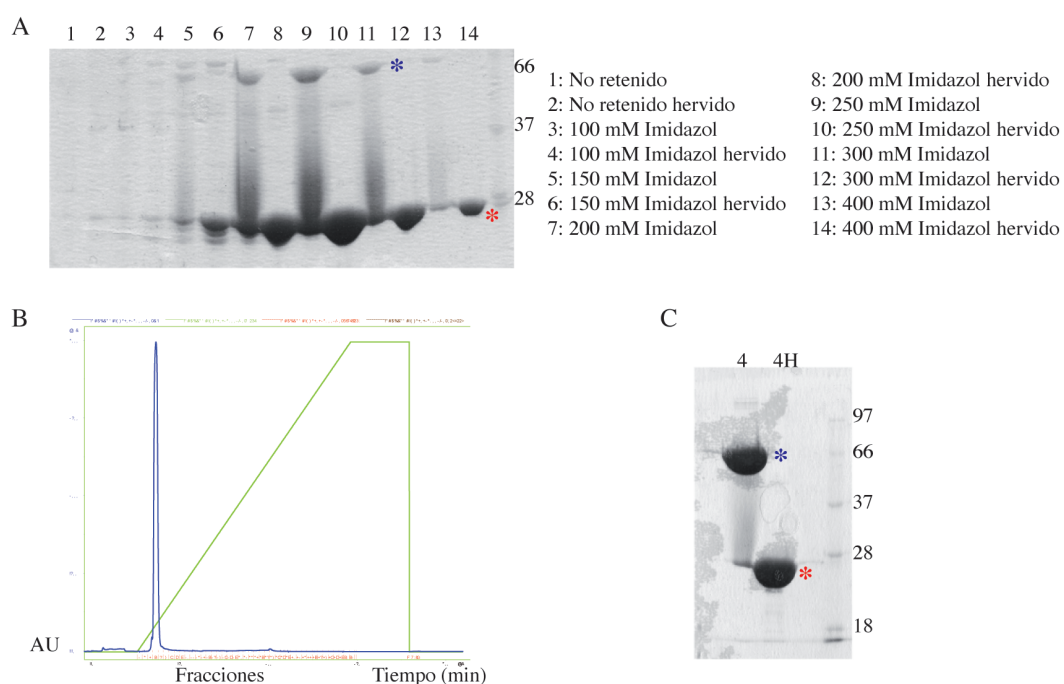


Figura 4.3: Purificación de gp17(371-553) con cola de histidinas. (A) Cromatografía de afinidad. Gp17(371-553) eluye a concentraciones superiores a 150 mM de imidazol. (B) Cromatografía de intercambio iónico. Se obtiene un único pico a 170 mM de cloruro sódico. (C) Gel de la fracción que contiene el pico. La proteína está pura y mantiene su estado trimérico. Los asteriscos azules indican las bandas correspondientes a los trímeros y los asteriscos rojos las bandas correspondientes a los monómeros.

4.1.3 Tripsinización de gp17

Para gp17 y el fragmento gp17(371-553) se hicieron proteólisis limitadas con el objetivo de obtener un fragmento más estable y por lo tanto incrementar sus posibilidades de cristalizar. Las muestras que se utilizaron eran de proteína obtenida

con cola de histidinas, y la tripsinización se hizo con fracciones obtenidas tras la purificación mediante cromatografía de afinidad. En ambos casos se comparó la muestra inicial con la proteína obtenida tras una tripsinización de 30 min y una de 2 h. Los resultados obtenidos tras la secuenciación amino terminal mostraron que gp17 se corta por acción de la tripsina después de la arginina 114 (Figura 4.4A), por lo que se estaría eliminando casi por completo la región amino terminal que interaccionaría con la cola del fago, siendo además una región en la que se predice mayor desorden. En el caso del fragmento gp17(371-553) se obtuvieron dos fragmentos, cortados tras la lisina 389 y la arginina 400 (Figura 4.4B). Según los datos que se habían tenido en cuenta para diseñar estos fragmentos, estos dos puntos deberían encontrarse en el segundo nódulo de la parte distal de la fibra.

Con las proteínas tripsinizadas se continuó la purificación con una cromatografía de intercambio iónico tal y como se realizaba en las purificaciones sin tripsinización. La proteína obtenida tanto para gp17 como para gp17(371-553) se concentró y se llevaron a cabo ensayos de cristalización. En ninguno de los dos casos se obtuvieron cristales.

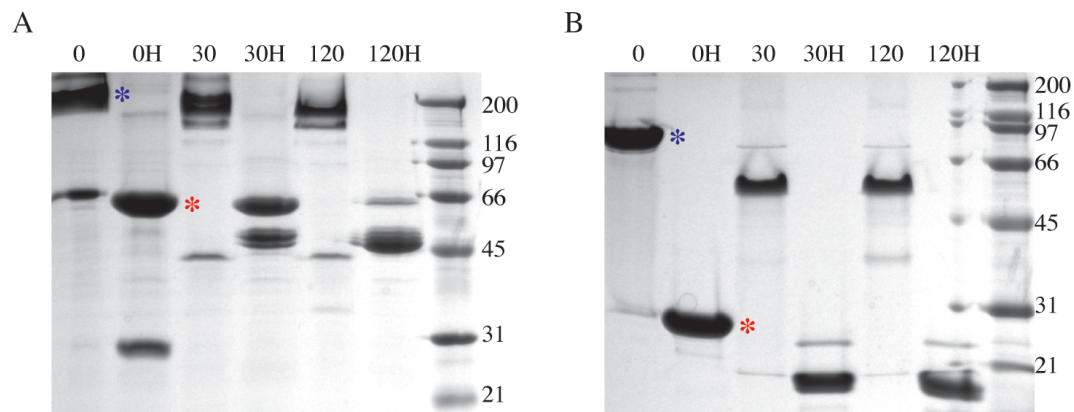


Figura 4.4: Tripsinización de gp17 y gp17(371-553). Gel de alícuotas tomadas durante la tripsinización, a tiempo 0, 30 y 120 min. En todos los casos se muestran fracciones no hervidas y hervidas. (A) gp17 (B) gp17(371-553). Los asteriscos azules indican los trímeros y los asteriscos rojos los monómeros, en ambos casos antes de la proteólisis.

4.1.4 Microscopía electrónica de las fibras de T7

Gp17 purificada se utilizó para obtener imágenes por microscopía electrónica y poder observar si estaba correctamente plegada. En las imágenes se observa que la proteína presenta la forma de L característica de las fibras laterales de los fagos (Figura 4.5), en la que se pueden diferenciar dos partes separadas por una bisagra o rodilla. Esta región bisagra forma un ángulo que varía entre las fibras depositadas

sobre la rejilla, lo que indica que es una región flexible. Una de las mitades de la fibra es más fina, y presenta un nódulo en su extremo. Esta parte se corresponde con la parte proximal de la fibra y el nódulo en el amino terminal podría corresponderse con la región de unión a la cola del fago. En la otra parte se observa la formación de regiones nodulares, correspondiéndose con los cuatro nódulos de la parte distal de la fibra tal y como se había descrito en el trabajo de Steven (Steven *et al.*, 1988) y que ya se han descrito en la página 71. La rejilla presenta un granulado que actúa como ruido a la hora de poder visualizar los cuatro nódulos, por lo que en algunas fibras es difícil diferenciarlos, pero el resultado indica claramente que la fibra se está plegando correctamente y con la forma que se esperaba.

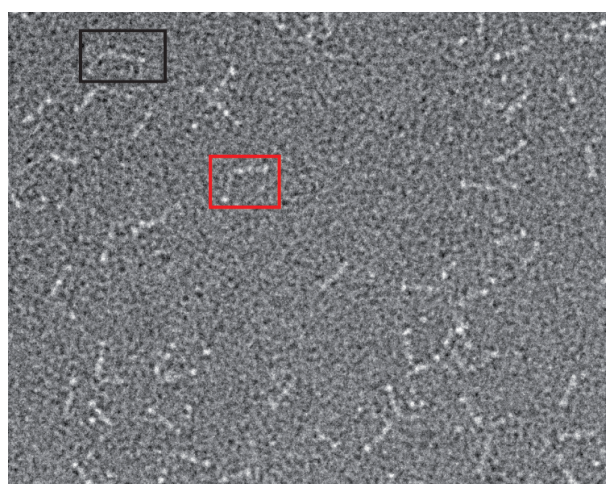


Figura 4.5: Imagen de fibras de T7 obtenida mediante microscopía electrónica. Aunque existe cierta heterogeneidad en la muestra, se puede observar como la mayor parte de las partículas tienen la morfología en forma de L característica. Se resaltan dos fibras, en el recuadro rojo se enmarca una fibra en la que se pueden diferenciar todos sus dominios y en el recuadro negro una con una fibra cuyo ángulo es mayor, mostrando la flexibilidad de la rodilla.

4.1.5 Cristalización del fragmento gp17(371-553)

Se obtuvieron cristales de este fragmento en gotas con 0,1 M Tris-HCl pH 8,5, 20% (m/v) PEG MME 2000 y 0,2 M trimetilamina-N-óxido (Figura 4.6A). Los cristales crecían en grupos que mostraban una morfología irregular en los bordes ya que crecían en forma de láminas superpuestas, pero estos clústeres se pudieron separar en la gota y aislar cristales más regulares. Éstos fueron congelados y medidos, y difractaron los rayos X hasta 1,9 Å (Figura 4.7B). De la misma forma, se obtuvo un cristal, en este caso con una morfología más regular, en 0,2 M citrato amónico pH 7 y 20% (m/v) PEG 3350, que difractó a 2 Å (Figura 4.6B). Los

primeros cristales pertenecían al grupo $P2_12_12_1$ y tenían una celdilla de 61,2 x 86 x 118,4; el segundo era un $C222_1$ con una celdilla 68,3 x 145,6 x 172,1. Las estadísticas obtenidas se encuentran en la Tabla 4.1. En ambos casos las predicciones indicaban un único trímero en la unidad asimétrica, con un contenido del 44 y 60% de solvente, respectivamente.

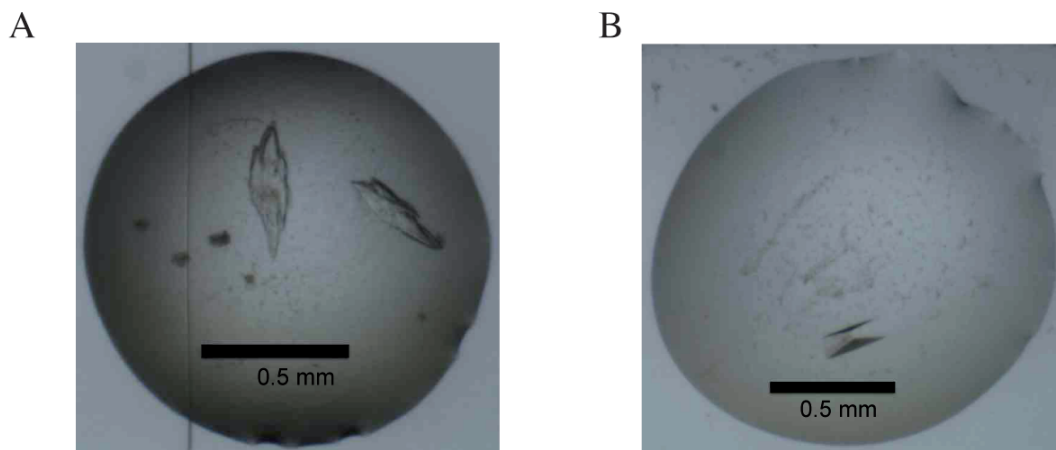


Figura 4.6: Cristales de gp17(371-553). (A) Cristales obtenidos en 0,1 M Tris-HCl pH 8,5, 20% (m/v) PEG MME 2000 y 0,2 M trimetilamina-N-óxido. (B) Cristal obtenido en 0,2 M citrato amónico pH 7 y 20% (m/v) PEG 3350.

A cristales del primer tipo se les añadió a la gota 1 μ l de condición de cristalización que previamente se había saturado con cloruro de metil mercurio durante varias horas, permitiendo que el átomo pesado penetrase en el cristal durante varias horas. Se recogieron datos a cuatro longitudes de onda de un cristal que difractó a 2,7 Å (Figura 4.7A), y que presentaba el mismo grupo espacial y unos parámetros de celdilla unidad similares, pero que habían incrementado su tamaño en un 3% en uno de los ejes (Tabla 4.2). Esta diferencia hacía imposible resolver la estructura mediante reemplazamiento isomorfo, pero se pudieron localizar las posiciones del mercurio mediante la técnica de dispersión anómala múltiple (MAD). Las estadísticas del proceso de faseado se encuentran reflejadas en la Tabla 4.4. Los datos obtenidos del cristal derivado presentan menor resolución que los nativos, probablemente debido a que, al añadir el mercurio, se altera el orden del cristal, disminuyendo la calidad de los datos recogidos.

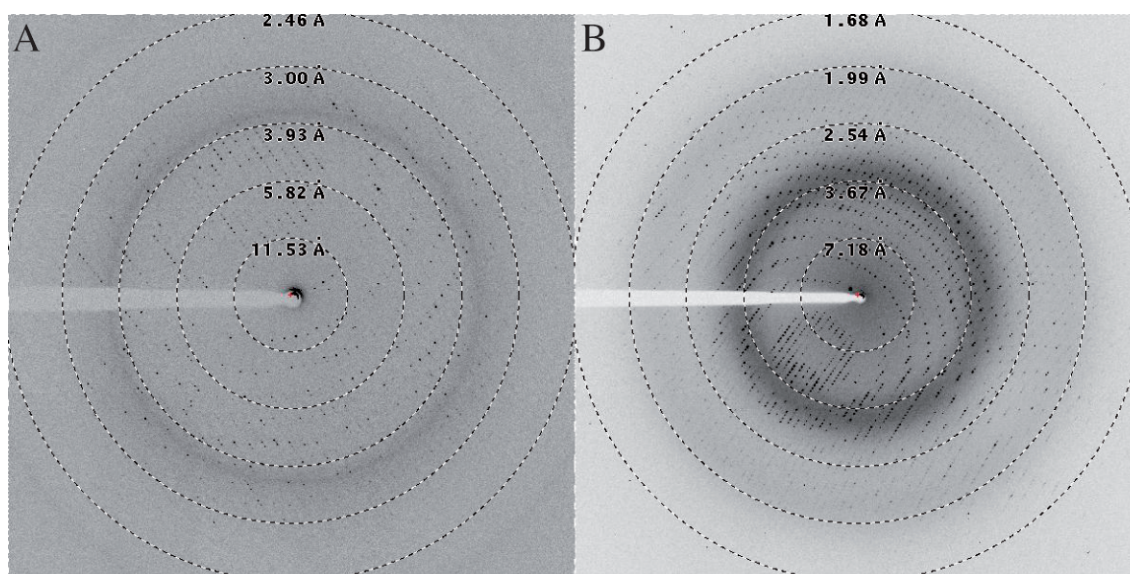


Figura 4.7: Patrones de difracción de cristales de gp17(371-553). Se representan diferentes anillos de resolución. (A) Cristal con mercurio utilizado para recoger datos a cuatro longitudes de onda diferentes. (B) Cristal nativo.

Tabla 4.1: Estadísticas obtenidas del escalado de los sets de datos nativos de gp17(371-553).

Grupo espacial	$P2_12_12_1$	$C222_1$
Parámetros de celdilla (a, b, c, Å)	61,3, 86,0, 118,4	68,3, 145,6, 172,1
Rango de resolución usado	20,0-1,9 (2,00-1,90)	20,0-2,0 (2,11-2,00)
Reflexiones	50056 (7224)	58236(8430)
Multiplicidad	3,7 (3,7)	3,7 (3,7)
Compleitud (%)	99,9 (100)	99,9 (100)
$\langle I/\sigma(I) \rangle$	6,3 (2,5)	6,0 (2,1)
R_{merge} (%)	9,3 (29,2)	9,7 (34,1)

Los valores entre paréntesis se corresponden con la capa de máxima resolución.

$$R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$$

Se localizaron nueve átomos de mercurio: seis están situados cerca de la Cys408 y Cys499 de las tres cadenas del trímero, dos cerca del Asp442 de las cadenas A y B y uno cerca de la His433 de la cadena A. El faseado y el aplanamiento de solvente se hizo con los datos a 2,7 Å, que permitieron construir automáticamente un modelo inicial con un total de 527 residuos. Este primer modelo se utilizó para un reemplazamiento molecular utilizando los datos nativos recogidos a 1,9 Å (Tabla 4.1) de un cristal del mismo tipo, construyéndose automáticamente un modelo de 528 residuos. Este modelo se completó manualmente y sobre él se realizó el ajuste y

Estructura de fibras de bacteriófagos

refinado, y se añadieron las moléculas de solvente observadas (3 moléculas de trimetilamina-N-óxido, una de PEG 400, una de Tris y 754 moléculas de agua).

Tabla 4.2: : Estadísticas obtenidas del escalado de los sets de datos utilizados para el MAD de gp17(371-553).

	Remoto de baja energía	Punto de inflexión	Pico	Remoto de alta energía
Longitud de onda (en Å)	1,0121	1,0070	1,0056	0,9919
Parámetros de la celdilla (a, b, c, Å)	61, 86, 122	61, 86, 122	61, 86, 122	61, 86, 122
Rango de resolución (Å)	20,0-2,7 (2,85-2,70)			
Reflexiones	18382 (2652)	18373 (2651)	18420 (2641)	18392 (2644)
Multiplicidad	7,3 (7,4)			
Compleitud	99,8 (100)			
$\langle I/\sigma(I) \rangle$	4,5 (1,6)	4,6 (1,7)	4,8 (2,1)	4,7 (2,1)
R_{merge} (%)	14,7 (46,2)	14,5 (44,8)	13,7 (35,6)	13,7 (35,9)

Los valores entre paréntesis se corresponden con la capa de máxima resolución.

$$R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$$

Tabla 4.3: Estadísticas del refinamiento de gp17(371-553).

Grupo espacial	$P2_12_12_1$	$C222_1$
Número de reflexiones usadas	47889(6859)	55922(8023)
Número de reflexiones usadas para el R_{free}	2111(293)	2273(322)
Factor R^{\S}	0,155 (0,187)	0,149(0,204)
R_{free}	0,200(0,218)	0,197(0,233)
Valor B medio/ Valor B Wilson (Å ²)	15,3/13,9	19,7/19,4
Estadísticas del Ramachandran [¶] (%)	97,2/99,4	97,3/99,5
Rmsd ^{**} (enlaces, Å/ángulos, °)	0,015/1,4	0,015/1,4

Los valores entre paréntesis se refieren a la capa de máxima resolución cuando es aplicable.

[§] $R = \sum ||F_{\text{obs}}(hkl)| - |F_{\text{calc}}(hkl)|| / \sum |F_{\text{obs}}(hkl)|$. [¶]Determinado con MOLPROBITY (Chen *et al.*, 2010). Los porcentajes indican los residuos en regiones favorecidas y permitidas respectivamente. ^{**}Obtenido con REFMAC (Murshudov *et al.*, 2011).

Tabla 4.4: Estadísticas del faseado de gp17(371-553).

Grupo espacial	P2 ₁ 2 ₁ 2 ₁			
Set de datos	Remoto de baja energía	Punto de inflexión	Pico	Remoto de alta energía
Longitud de onda (en Å)	1,0121	1,0070	1,0056	0,9919
Reflexiones (acéntricas/céntricas)	15813/2352	15814/2353	15770/2344	15812/2348
Átomos pesados	9 Hg ²⁺			
Coefficiente de correlación (todos/débiles)	55,88/33,63			
Figura de mérito de Patterson	20,42			
Coefficiente de correlación (E)	0,370			
R-Cullis [§]				
Isomorfo (acéntrico/céntrico)	0,819/0,842	-/-	0,795/0,786	0,630/0,640
Anómalo (acéntrico)	0,955	0,875	0,803	0,800
Poder de faseado [§]				
Isomorfo (acéntrico/céntrico)	0,579/0,462	-/-	0,492/0,389	1,033/0,784
Anómalo (acéntrico)	0,416	0,768	1,123	1,111
Figura de mérito (acéntrica/céntrica) [§]	0,528/0,322			
Aplanamiento de solvente (27 ciclos con 43,1% de solvente) [¶]				
Factor R (antes/después de la modificación de densidad)	0,4370/0,2520			
Correlación media de E ² (antes/después de la modificación de densidad) [¶]	0,4225/0,7900			
Puntuación de la mano = correlación de E ² /contraste (original/inversa)	0,6710/0,2265			

Encontrados con SHELXD (Sheldrick, 2008). [§]Calculado con SHARP (Vonrhein *et al.*, 2007). [¶]Según SOLOMON (Abrahams y Leslie, 1996).

Paralelamente se utilizó el modelo intermedio con 527 aminoácidos para resolver la estructura de la otra forma cristalina que difractaba a 2 Å, siguiendo un proceso similar (Tabla 4.1). En este caso las moléculas de solvente que presentaban densidad eran una molécula de carbonato y 910 moléculas de agua. Las estadísticas obtenidas del refinamiento de ambas formas cristalinas se encuentran reflejadas en la Tabla 4.3.

Durante la validación del modelo se observó que hay un aminoácido, la Gly522, que en ambos casos se encuentra en regiones poco favorecidas del gráfico de

Ramachandran, pero ya que este residuo se sitúa en la densidad electrónica observada, es probable que dicha conformación esté forzada por el resto de la estructura.

4.1.6 Descripción de la estructura

La estructura se divide en dos regiones claramente diferenciadas (Figura 4.8): el dominio pirámide y el dominio punta. El dominio pirámide está formado por cadenas β , abarca desde el comienzo del fragmento expresado (Gly371) hasta el Trp454 y se correspondería con los nódulos segundo y tercero de la microscopía electrónica. A continuación de este dominio se encuentran tres hélices α , una por cada monómero, conectando el dominio pirámide con el dominio punta. El dominio punta es más globular, correspondiéndose con el último nódulo que se observaba en las imágenes de microscopía, en las que se podía apreciar que tiene un tamaño mayor que los dos anteriores. Este dominio se extiende desde el final de la hélice α , aproximadamente desde la Ala465, hasta el extremo carboxilo terminal de la proteína (Glu553) y está formado también por cadenas β .

Al haber obtenido dos formas cristalinas diferentes, se superpusieron ambas estructuras para poder determinar el grado de diferencia entre ellas. Para ello se excluyeron los dos primeros residuos (371 y 372) dado que presentaban diferentes conformaciones, parcialmente debido al empaquetamiento de la proteína en el cristal. Comparando las estructuras completas, la diferencia (*rmsd*) entre la cadena principal de ambas es de 1,2 Å. Si se comparan los monómeros se observa una mayor desviación en el giro β entre los residuos 390 a 393 de una de las cadenas, pero al analizar el empaquetamiento se ha observado que esta desviación se debe a un contacto cristalino presente en la forma C222₁. Sorprendentemente, al analizar la diferencia de cada uno de los dominios por separado, se han obtenido valores de 0,2 Å para el dominio punta y de 0,5 Å para el dominio pirámide. Esto indica claramente que la región que conecta ambos dominios es la más flexible, lo que permite que exista cierto movimiento de los dos dominios entre las dos formas cristalinas analizadas. A su vez, esta flexibilidad permitiría el movimiento del dominio punta, más globular, y podría ser importante en el proceso de infección, facilitando el acceso del dominio punta a los receptores de la superficie celular sin que esto modifique el plegamiento o la posición del resto de la fibra.

Para nombrar las diferentes estructuras presentes en la proteína, se ha denominado la pequeña hélice α como A, seguida por las cadenas β del dominio punta (B-I). Para el dominio pirámide, la última cadena β antes de la hélice α se ha denominado Z, y se ha continuado de forma decreciente hacia el extremo amino terminal del fragmento, siendo la última cadena β descrita la R (Figura 4.8A).

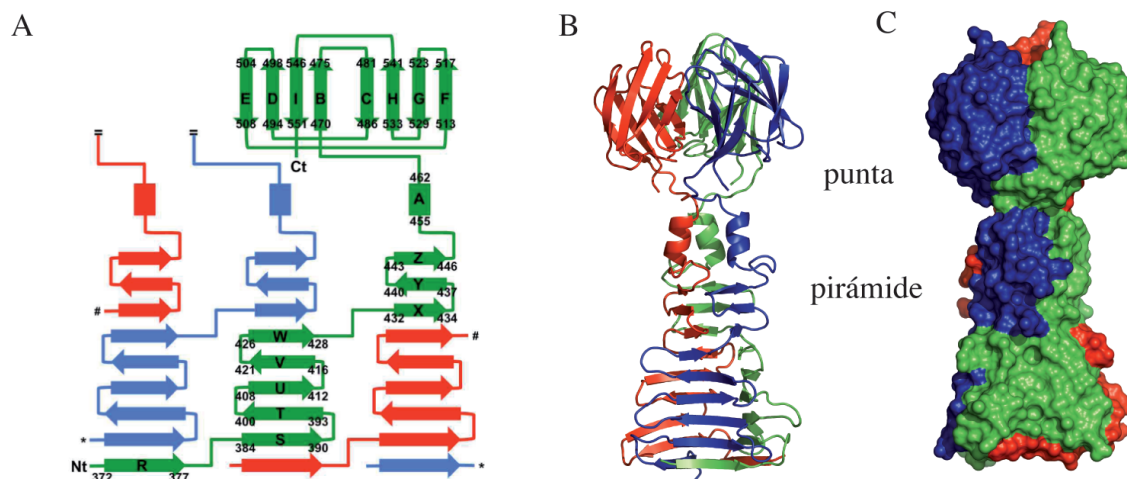


Figura 4.8: Estructura de gp17(371-553). (A) Topología de gp17(371-553). (B) Representación en la que se observan las cadenas β , y las pequeñas hélices α separando los dos dominios de la proteína. (C) Representación de la superficie de la proteína.

4.1.7 El dominio pirámide

Las láminas β que forman el dominio pirámide dotan a las paredes de esta parte de la fibra de una forma cóncava. Cada uno de los lados de la pirámide está formado por una lámina β de carácter mixto y está formada por cadenas β de los tres monómeros. La longitud de las cadenas β es variable, pero se observa una tendencia a disminuir la longitud hacia la punta de la pirámide, siendo la más larga la T con ocho aminoácidos y las más cortas la W y la X, ambas con tres aminoácidos (Figura 4.8B). En las láminas β la disposición de las cadenas β entre sí de un mismo monómero es siempre antiparalela, pero cuando es entre dos cadenas β de monómeros diferentes es paralela. Así, cada lámina β está formada por la cadena R del primer monómero, las cadenas S a W del segundo y X a Z del tercero, generando un giro dextrógiro (Figura 4.8B).

Si se analiza la composición en aminoácidos de estas láminas β , se observa que aproximadamente la mitad de los residuos que las forman son residuos hidrofóbicos. Entre ellos destacan la valina y la leucina, que están presentes en un 13 y 11%

respectivamente. Las cadenas laterales de estos residuos tienden a estar hacia el interior de la fibra, creando así un núcleo hidrofóbico, especialmente en el caso de la leucina. En menor proporción se encuentran residuos aromáticos, pero estos al igual que las leucinas, muestran una tendencia a orientar sus cadenas laterales hacia el interior de la estructura. Por otra parte, aunque su abundancia relativa no es tan destacable, sí se observa una preferencia por posiciones en las cadenas β de los aminoácidos ácidos o básicos, situándose casi siempre en los extremos de las mismas, siendo también mayoritarios en los lazos que conectan las cadenas β entre sí (Figura 4.9).

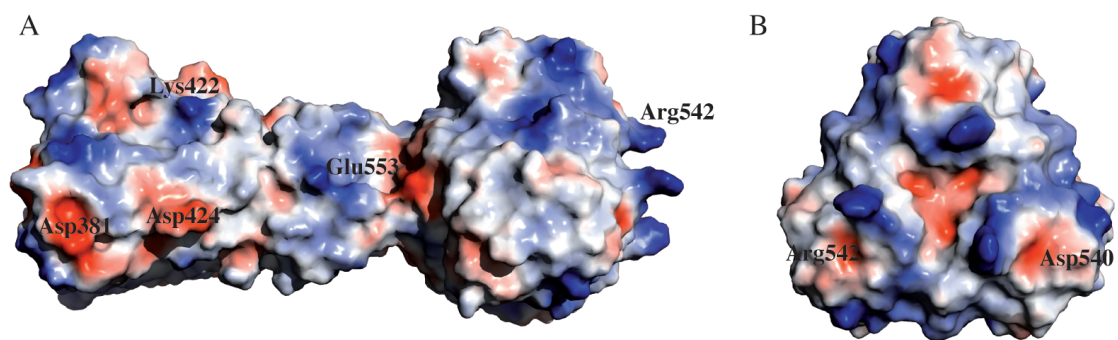


Figura 4.9: Representación de la superficie de la gp17(371-553) en función de la carga de sus aminoácidos. En azul aparecen representados los aminoácidos básicos y en rojo los ácidos. (A) Vista lateral. (B) Vista frontal.

Además, como previamente se habían realizado experimentos de tripsinización con este fragmento y se habían identificado los puntos de corte, se localizaron los dos puntos en la estructura, situándose ambos al final de dos cadenas β consecutivas, uno de ellos al final de la S, siendo el corte en la Lys389 y terminando ésta en el 390, y el segundo justo en el último aminoácido, la Arg400, de la cadena T (Figura 4.8A). Estas últimas cadenas β situadas hacia el extremo amino terminal podrían desplegarse con mayor facilidad en el fragmento expresado, que además poseía una cola de histidinas, facilitando el acceso a la tripsina. Este mismo desplegamiento sería el que en las electroforesis en las que las muestras no están hervidas da lugar a varias bandas, que se corresponderían con múltiples conformaciones posibles en el extremo amino terminal del fragmento gp17(371-553).

En el extremo de este dominio pirámide se encuentran tres hélices α pequeñas que, al igual que las láminas β en el dominio pirámide, también forman un núcleo

hidrofóbico en el que los aminoácidos polares apuntan hacia el exterior, siendo los más hidrofóbicos los que se encuentran hacia el interior, destacando por su orientación la Leu459.

4.1.8 El dominio punta

Las hélices α del extremo del dominio pirámide se conectan al dominio punta a través de un lazo. La composición de este lazo diverge frente a lo previamente observado ya que en él se encuentran residuos tanto polares como apolares, sin formar por lo tanto una región hidrofóbica, e incluso permitiendo la entrada a moléculas de solvente en la forma cristalina. A su vez, el hecho de que pequeñas moléculas de solvente puedan interaccionar con esta región, puede apoyar la teoría previamente explicada sobre su flexibilidad y sus posibles implicaciones en el proceso de infección.

En este dominio los tres monómeros no están entrelazados, estando formado cada uno de ellos por un sándwich β , compuesto por las diferentes cadenas β de ese monómero intercaladas. Una de las partes está formada por las cadenas EDIB y la otra por CHGF (Figura 4.8A), quedando las cadenas centrales B y C apuntando hacia la parte externa y las cadenas E y F hacia el interior del trímero, más próximas a las respectivas cadenas de los otros dos monómeros. De nuevo, al igual que en el dominio pirámide, la mayor parte de los residuos de este dominio son hidrofóbicos, y tienen sus cadenas laterales orientadas hacia el interior, destacando principalmente la presencia de valina (10%), isoleucina (14%) y fenilalanina (10%).

Los lazos CD, EF y GH se sitúan hacia la región que conecta con el dominio pirámide. En cambio, los lazos BC, DE, FG y HI se sitúan hacia el exterior. Además, los lazos DE y FG conectan cadenas de la misma lámina β , mientras que los BC y HI de diferentes. Los lazos que quedan expuestos hacia el exterior son hidrofóbicos e hidrofílicos en la misma proporción, y no hay aminoácidos aromáticos. En cuanto a los aminoácidos con carga, solamente hay una arginina en el lazo HI y un ácido aspártico en el lazo FG. Respecto a su disposición, la mayor parte de las cadenas laterales quedan expuestas, destacando la Arg542, cuya cadena lateral forma una protuberancia en la superficie de la estructura (Figura 4.9). De forma contraria, el Asp520, aunque tiene la cadena lateral expuesta, está orientada hacia el sándwich β formado por el monómero contiguo. Analizando las distancias entre los átomos de este dominio, se han localizado tres pares de aminoácidos cuyos átomos podrían

estar formando enlaces de hidrógeno o puentes salinos entre diferentes monómeros. El primero de ellos incluye este Asp520, cuya orientación le permite formar enlaces con la Arg525 de la cadena G del monómero contiguo. Los otros dos pares, Glu553-Lys468 y Arg508-Glu551 se encuentran en los lazos de la parte inferior del dominio punta, ambos estabilizando el extremo carboxilo terminal, orientándolo hacia el interior de la base del dominio punta, y protegiéndolo de la posible degradación por proteasas. Estas interacciones se producen entre las tres cadenas, de forma que si nos fijamos en los aminoácidos Glu551 y Glu553 de una cadena, estos interaccionarían con la Lys468 de la segunda cadena y la Arg508 de la tercera.

4.1.9 Estabilidad y plegamiento de gp17(371-553)

Los lazos que conectan las cadenas β entre ellas o con la pequeña hélice α , tanto en el dominio punta como en el pirámide, muestran una densidad clara, y muchos se encuentran estabilizados mediante interacciones con otras cadenas, pudiendo ser mediante puentes salinos, de hidrógeno, fuerzas de van der Waals o mediante las moléculas de solvente ordenadas. Se ha calculado el número de enlaces que podrían formarse, siendo 90 enlaces de hidrógeno y 8 puentes salinos en el dominio pirámide y 27 enlaces de hidrógeno y 30 puentes salinos en el dominio punta.

Otro de los factores que influyen en la estabilidad de la estructura es la superficie expuesta. En cada monómero, se ha calculado que la superficie expuesta sería de $12 \times 10^3 \text{ \AA}^2$, pero al formarse el trímero más del 40% queda enterrada ($5,2 \times 10^3 \text{ \AA}^2$), y la energía necesaria para su disociación sería de 125 kcal/mol. Analizando los dos dominios por separado, en el dominio pirámide la superficie expuesta de cada monómero es de $7,4 \times 10^3 \text{ \AA}^2$, de la que aproximadamente la mitad ($3,6 \times 10^3 \text{ \AA}^2$) queda enterrada en el trímero; y en el dominio punta la superficie expuesta es de $4,7 \times 10^3 \text{ \AA}^2$, de los cuales el 30% ($1,4 \times 10^3 \text{ \AA}^2$) no están expuestos en el trímero. No se han observado puentes disulfuro y, como sólo hay dos cisteínas en este fragmento que no están próximas y otra más en el resto de la proteína, no es probable que existan.

En el dominio punta los tres monómeros tienen un plegamiento relativamente independiente y no se entrecruzan. Si se suma esto al hecho de que la superficie enterrada es menor y la energía de disociación también es menor, podría interpretarse que este dominio es el punto de inicio del plegamiento de la fibra. De forma independiente se formarían los tres barriles β monoméricos que interaccionarían para formar el trímero. El resto de la proteína continuaría plegándose ya de forma

trimérica con un "efecto cremallera". Este tipo de plegamiento es similar al propuesto previamente para el adenovirus (Mitraki *et al.*, 1999), y es congruente con los ensayos realizados con mutantes *amber* de gp17, en los que se observó que es necesario que la proteína se exprese completa para que trimerice y se acople a la cola del fago (Kemp *et al.*, 2005). También hay que destacar que gp17 no parece requerir la presencia de chaperonas para su plegamiento, como en el caso de las fibras de otros fagos (como por ejemplo T4, que para el correcto plegamiento de las proteínas de sus fibras necesita la chaperona gp57 y, en el caso de gp37 que forma el extremo de la parte distal de la fibra larga, también la chaperona gp38) (Bartual *et al.*, 2010a). La necesidad de chaperonas para el proceso de trimerización también ha sido descrita en el caso del fago K1F (Schulz *et al.*, 2010a) que presenta una chaperona intramolecular similar a la presente en la fibra de T5, cuya función será descrita más adelante en este trabajo.

4.1.10 Unión a receptor

La unión inicial y reversible del bacteriófago T7 a la bacteria hospedadora se realiza a través de la interacción de las seis fibras de con los LPS. Tras este primer reconocimiento podría llevarse a cabo una unión secundaria e irreversible a través de la cola, pero hasta el momento no se ha descrito un receptor secundario. El papel exacto de la unión al receptor (o receptores) a la hora de determinar el rango de hospedador todavía no está suficientemente estudiado. Se ha propuesto que gp17 podría simplemente mantener al fago en el entorno de la célula, permitiendo una difusión bidimensional, que facilitaría la unión irreversible (Hu *et al.*, 2013).

Por otra parte, se han llevado a cabo estudios en los que se ha modificado mediante evolución dirigida el rango de hospedador mediante la mutación de un solo aminoácido: cambiando el Asp520 a Glu, el fago no puede infectar *E. coli* B, y cambiando la Val544 a Ala, se evita la infección en *E. coli* K-12 (Heineman *et al.*, 2008). También se ha identificado en el fago ϕ A1122 un mutante espontáneo en el aminoácido 523, que es el equivalente al 518 en T7, que modifica el rango de hospedador (García *et al.*, 2003). Estos tres aminoácidos que han sido descritos como relacionados con la interacción con el hospedador se encuentran en lazos expuestos en el extremo más distal del dominio punta. Además, si se comparan las secuencias de la fibra de T7 con las de otros fagos similares (Figura 4.10A), se puede observar que la mayor parte de las diferencias se encuentran en estos lazos en el extremo del

Estructura de fibras de bacteriófagos

dominio punta (Figura 4.10B y C), por lo que se puede suponer que dichos lazos están implicados en el proceso de unión reversible al hospedador, interaccionando con regiones de los LPS que presentan diferente grado de exposición en función de la estructura que tengan en la cepa utilizada.

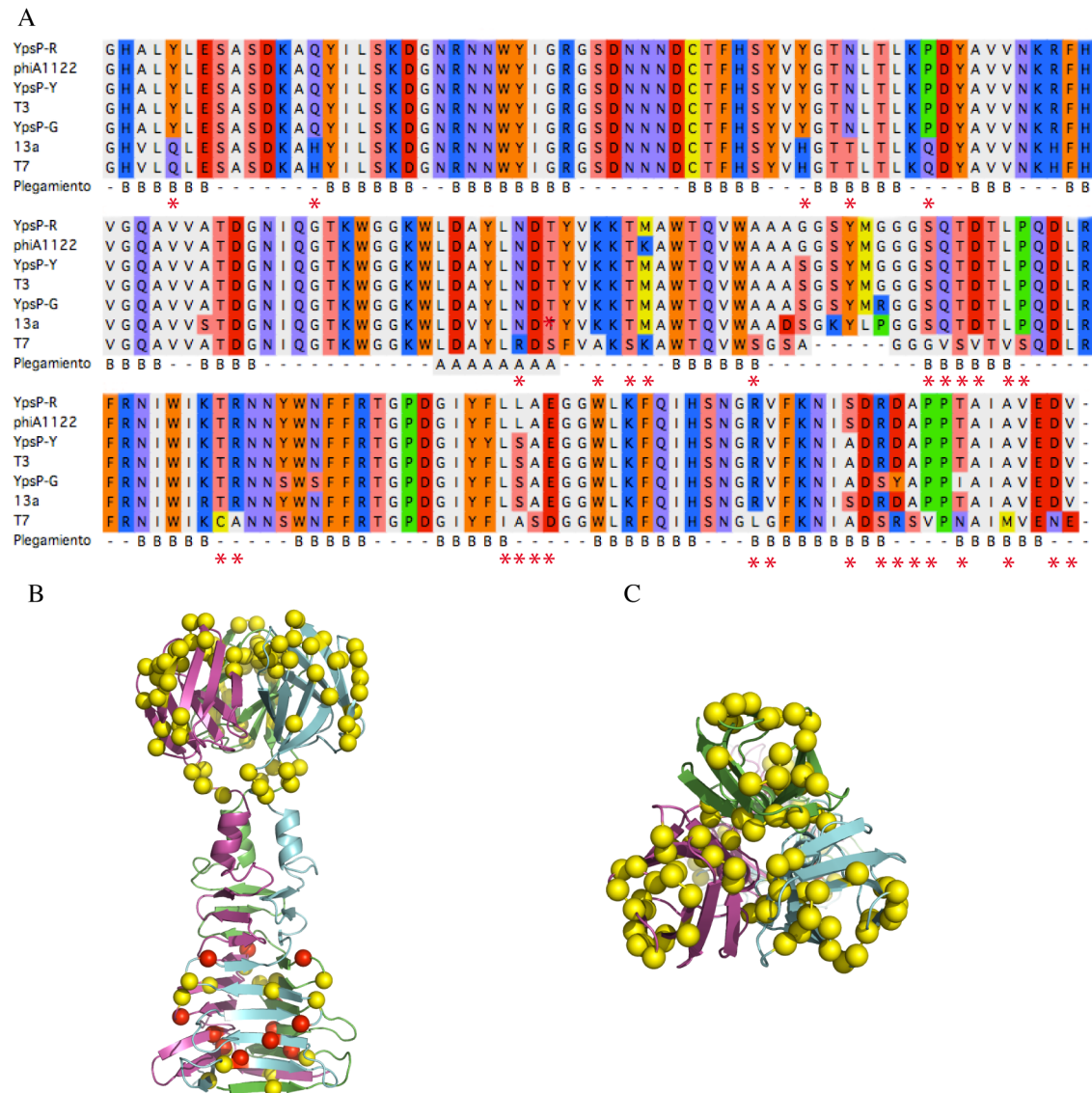


Figura 4.10: Conservación de secuencia en gp17(371-553). (A) Alineamiento de la secuencia de gp17(371-553) con las secuencias de las fibras de otros fagos. En la última fila se representa el plegamiento (A para hélices α y B para cadenas β). Se destacan con un asterisco rojo aquellos que en gp17 son diferentes de al menos dos fibras. (B) Vista lateral de gp17(371-553). (C) Vista frontal de gp17(371-553). En B y C se destacan los carbonos α de los aminoácidos resaltados en el apartado A en amarillo. En rojo se destacan los carbonos α de los aminoácidos que interaccionan con moléculas de solvente.

También existe otra región que podría estar implicada en la unión al receptor: las láminas β del dominio pirámide. Esta región forma una concavidad y en ella se concentran varios residuos aromáticos. Además, en ambas formas cristalinas, en esta región se observa la interacción con moléculas de solvente que no son aguas, siendo entre la trimetilamina-N-óxido y los aminoácidos Tyr385, Tyr397 y Tyr413 en la forma $P2_12_12_1$ y entre un carbonato y la Tyr425 en dos de los tres monómeros en la forma $C222_1$ (Figura 4.10B y C). En cualquier caso, hay que tener en cuenta que no existen estudios bioquímicos o mutacionales sobre esta región, que serían necesarios para afirmar que dichos aminoácidos pueden tener un papel relevante en la unión a los LPS.

4.2 Ensayos de inhibición de la infección por T7

Mediante una serie de ensayos de inhibición se pretendía determinar si era posible que gp17 bloquease la infección de *E. coli* por T7 o si era posible inhibir la unión de las fibras del fago con lipopolisacáridos.

En el ensayo 1 se determinó la dilución necesaria para la muestra de fagos con la que se iban a hacer los sucesivos ensayos. Con una dilución 10^{-10} se obtenía un número de placas de lisis de entre 50 y 100, así que se seleccionó esta dilución. De forma paralela se determinó que la diferencia en la capacidad de infección del fago en células de la cepa BL21 o BL21(DE3) no era notable. Una vez seleccionada la cepa (BL21) y la concentración de fagos (10^{-10}), se analizó si la presencia de gp17 interfería en la infección. Para ello, en el ensayo 2, se determinó que el número de placas de lisis obtenidas tras añadir 1,5 mg de BSA o de gp17 no variaba notablemente, por lo que gp17 no parecía estar inhibiendo la infección. En este segundo ensayo se usó BSA como control, ya que no debería interferir en la unión entre el fago y las células. Los resultados de este ensayo se corroboraron en el ensayo 3, repitiendo el mismo experimento de forma independiente.

En los dos ensayos anteriores se observaba que había una pequeña variación en la media del número de placas de lisis. Para poder asegurar que los resultados procediesen de análisis totalmente independientes, se realizó el mismo experimento utilizando dos muestras de gp17(371-553) y una tercera de la misma proteína, pero tripsinizada (ensayo 4). Se observó que con la primera muestra disminuía el número de placas de lisis, pero no con las otras dos, por lo que se valoró la posibilidad de que alguna otra molécula que permaneciese en la muestra pudiese estar inhibiendo la

formación de placas de lisis. Para comprobar si esta inhibición se debía a elementos residuales de la purificación, se añadieron a la infección lisados bacterianos (ensayo 5) de cultivos en los que se había sobreexpresado gp17 con y sin cola de histidinas (ya que este factor también podría interferir en la unión). No se observó ningún efecto en el número de placas de lisis.

Cuando se llevan a cabo purificaciones de fibras virales, uno de los contaminantes más habituales podrían ser los propios LPS a los que se unen estas fibras, por lo que se valoró que fuesen los LPS los que estuviesen inhibiendo la infección bloqueando la fibra, y no la proteína bloqueando el sitio de unión en la bacteria. En el ensayo 6 se comprobó el efecto de dos tipos de LPS diferentes, sin que disminuyese el número de placas de lisis. El ensayo se repitió de forma independiente (ensayo 7), obteniendo los mismos resultados.

Para poder descartar totalmente el efecto de la proteína, se hizo un último ensayo (ensayo 8) incubando las muestras con gp17(371-553) con y sin cola de histidinas, y también se incluyeron los dos LPS que se habían utilizado en los dos ensayos anteriores. Además, en este ensayo se incluyó un nuevo LPS, el EH100, y se observó que este último inhibía completamente la formación de placas de lisis.

Tras haber identificado el LPS EH100 como inhibidor de la formación de placas de lisis, se realizaron dos ensayos más (ensayos 9 y 10) para calcular la concentración de EH100 necesaria para inhibir la formación de placas de lisis. Se calculó que cuando la cantidad de LPS EH100 es menor de 1 µg, el aumento del número de placas de lisis es directamente proporcional a la disminución de la cantidad de LPS, existiendo entre ambos valores la siguiente relación:

$$\text{Número de placas de lisis} = -24 \times (\mu\text{g de LPS}) + 41$$

Los datos se ajustan a dicha relación con un coeficiente de correlación $R^2 = 0,995$ (Figura 4.11). El número de placas de lisis obtenidas en cada uno de los ensayos aparece reflejado en la Tabla 4.5.

Según los cálculos anteriores, la cantidad de LPS necesaria para reducir a la mitad el número de placas de lisis obtenidas es de 0,84 µg de LPS. Dicho de otra forma, 10 ng de LPS EH100 inhiben aproximadamente 20 fagos en las condiciones en las que se realizó el ensayo, en las que se obtenían 40-50 pfu y la concentración del lipopolisacárido era de aproximadamente 10 ng/µl en la reacción.

Estos resultados muestran que, aunque la presencia de gp17 no es suficiente para inhibir la infección por el fago T7, si se añade el receptor a la reacción, éste sí puede interaccionar con el fago, evitando que se una a las bacterias. En el ensayo 4 con la primera muestra de proteína sí se obtuvo una disminución significativa del número de placas de lisis al añadir gp17, pero al ser un caso aislado en una purificación puntual, podría deberse a una contaminación de lipopolisacáridos presentes en la muestra, ya que en ningún otro experimento se observó esta inhibición.

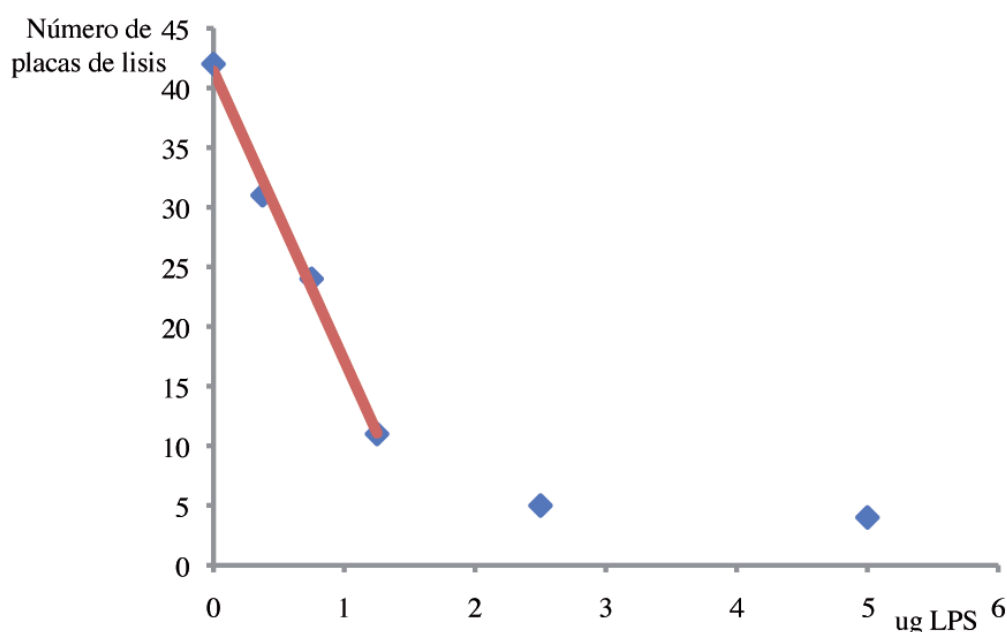


Figura 4.11: Inhibición de la formación de placas de lisis mediante la unión de LPS. Los puntos azules representan el número de placas de lisis observado en función de la cantidad de LPS. La línea roja muestra la correlación lineal entre el aumento de LPS añadidos y la disminución de placas de lisis observadas.

Durante el proceso de infección es probable que la unión de las fibras sea transitoria, de forma que el fago no se mantenga estable sobre la célula hasta que todas las fibras estén unidas. Recientemente se ha propuesto un modelo mediante el cual se produciría una serie de movimientos en los cuales las fibras se irían uniendo y soltando mientras el fago se desplaza buscando el lugar idóneo para la infección y que, una vez encontrado, se unirían el resto de fibras para facilitar la interacción de la cola y desencadenar la infección (Hu *et al.*, 2013). Así pues, la unión de las fibras de forma individual sería muy débil, y por ello no se observa diferencia en la infección cuando se hacen ensayos de competición entre las fibras y el fago completo, que sí

Estructura de fibras de bacteriófagos

puede formar uniones más estables al unir el resto de fibras. También hay que tener en cuenta que los LPS en solución que se han utilizado para los ensayos de inhibición forman agregados que, para las fibras, mimetizan las células.

Tabla 4.5: Ensayos de inhibición de la infección del fago T7.

Ensayo 1				
Cepa		10^{-9}	10^{-10}	10^{-11}
BL21		>100	58	7
BL21(DE3)		>100	66	11
Ensayo 2				
Proteína	Placa 1	Placa 2	Placa 3	Media
BSA	77	64	61	67±9
gp17(371-553)	68	62	61	64±4
Ensayo 3				
Proteína	Placa 1	Placa 2		Media
BSA	46	36		41±7
gp17(371-553)	38	38		38
Ensayo 4				
Proteína	Placa 1	Placa 2		Media
BSA	154	186		170±23
gp17(371-553)1	137	105		121±23
gp17(371-553)2	153	171		162±13
gp17(371-553)Tr	178	-		178
Ensayo 5				
Lisado	Placa 1	Placa 2		Media
600 µl no cola	57	47		52±7
300 µl no cola	42	44		43±2
150 µl no cola	43	43		43
100 µl con cola	43	36		40±5
50 µl con cola	30	44		37±10
25 µl con cola	36	45		41±6
Ensayo 6				
LPS	Placa 1	Placa 2		Media
O55	57	47		52±7
O111	42	44		43±1
Ensayo 7				
Bacterias	27	29	28	28±1
20 µl O111	38	28	40	35±6
20 µl O55	32	28	41	34±7

Ensayo 8							
Proteína/LPS			Placa 1		Placa 2		Media
TMS			46		62		54±11
gp17(371-553) con cola			63		49		56±10
gp17(371-553) sin cola			44		58		51±10
EH100			0		0		0
Ensayo 9							
EH100	200 µg	100 µg	10 µg	1 µg	100 ng	10 ng	1 ng
Placa 1	0	0	0	13	52	71	58
Ensayo 10							
EH100	5 µg	2,5 µg	1,25 µg	0,75 µg	0,375 µg	TMS	
Placa 1	4	5	11	24	31	42	

Se indican en cada caso el número de placas de lisis obtenidas.

Los LPS que han mostrado una clara inhibición de la infección de T7, los EH100, son de un mutante Ra (carecen sólo del antígeno O), mostrando una estructura de la endotoxina similar a la de las células *E. coli* K-12. En el caso de las células utilizadas para la infección, la cepa *E. coli* BL21 es derivada de *E. coli* B, que en la estructura de sus LPS carece de la parte más externa si la comparamos con K-12. Durante los experimentos también se hicieron pruebas con otros dos LPS que presentaban antígeno O, siendo todos los resultados negativos y apoyando la teoría de que es necesario que el núcleo de los LPS esté suficientemente expuesto para que se pueda producir la infección (Qimron *et al.*, 2006). Si se comparan los resultados obtenidos con los estudios previamente publicados sobre infecciones de T7 en mutantes de LPS (Garcia *et al.*, 2003), se confirma que es suficiente la presencia del núcleo de heptosas más la primera glucosa para la infección (el caso de las células B), sin que se pueda determinar si las dos glucosas más la heptosa restantes en las células K-12 tienen algún efecto en la infección, aunque sí podemos comprobar que la presencia del antígeno O inhibe la infección.

En un futuro, para continuar investigando la unión, sería necesario realizar las diferentes infecciones utilizando una batería de fagos con la fibra mutada, así como utilizar, si fuese posible, otros LPS que presenten diferencias en la estructura del núcleo. Ahora que se ha resuelto la estructura y teniendo en cuenta los trabajos previos sobre mutantes puntuales, se podrían construir mutantes dirigidos de la fibra en las regiones que es más probable que estén interaccionando con el receptor, especialmente los lazos del dominio punta. Además, comparando los resultados de la

infección en células con diferente estructura en sus LPS, se podría comenzar a clarificar los puntos exactos de interacción entre los LPS y la fibra de T7.

4.3 Comparación de gp17 con otras proteínas

Tras realizar la búsqueda de secuencias similares, se obtuvieron las proteínas cuyas secuencias presentaban una identidad máxima superior al 75% con gp17 y cuyo valor E (Expect value, es una medida del ruido de fondo, y describe el número de resultados iguales que se obtendrían por azar) era igual a 0, que se presentan en la Tabla 4.6. No se consideró un número mayor de secuencias dada la gran diferencia en el porcentaje de identidad (63% el siguiente resultado) y debido a que la similitud se localizaba únicamente en el extremo amino terminal, y por lo tanto no proporcionarían información sobre diferencias que pudieran intervenir en la especificidad de unión al receptor.

Al alinear las secuencias se observa que la región más variable es la carboxilo terminal, que es la que interviene en la unión al hospedador. Entre las diferencias más destacables se encuentra la delección de 5 aminoácidos a partir del 480 en el caso de gp17 frente a las otras secuencias. Esta delección podría estar implicada en el cambio de rango de hospedador, ya que estos aminoácidos se encuentran en el lazo BC del dominio punta, expuesto hacia el extremo distal. El resto de diferencias se muestran en la Figura 4.12.

Tabla 4.6: Fibras que presentan homología de secuencia con gp17 de T7.

Proteína	Nº acceso	Identidad con gp17 de T7
gp17 del fago 13a	YP_002003979	90%
Fibra de la cola del fago de <i>Yersinia</i> YpsP-G	AFK13534	87%
Fibra de la cola del fago T3	NP_523342	87%
Fibra de la cola del fago ϕ A1122	NP_848304	87%
Fibra de la cola del fago de <i>Yersinia</i> YpsP-Y	AFK13385	86%
Fibra de la cola del fago de <i>Yersinia</i> YpsP-R	AFK13438	86%

Se incluye el código de acceso de la base de datos de proteínas del NCBI y porcentaje de identidad.

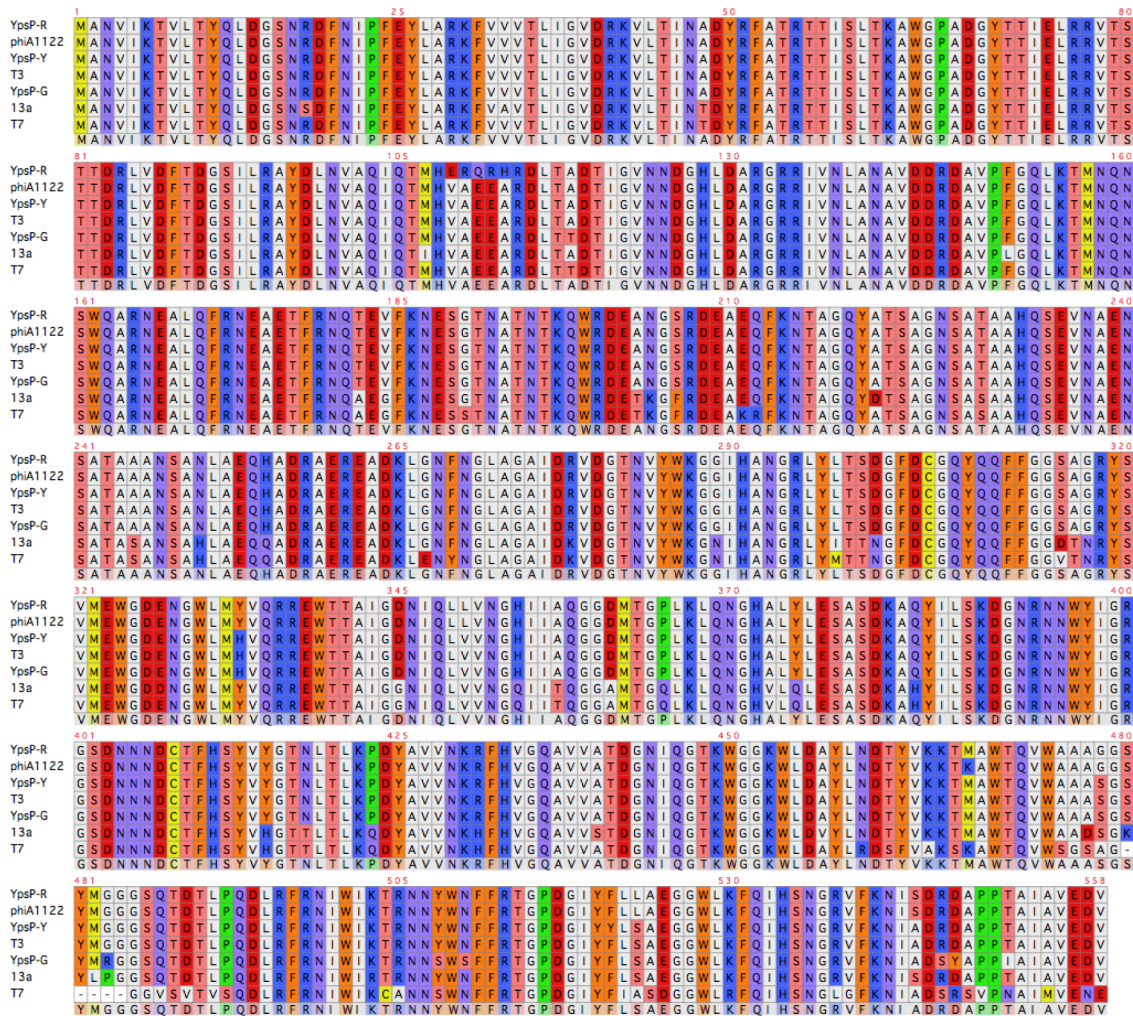


Figura 4.12: Alineamiento de las secuencias completas de las fibras de YpsP-R, phiA1122, YpsP-Y, T3, YpsP-G, 13a y T7.

Entre las secuencias descartadas para la comparación, se han encontrado fibras de fagos que infectan otros géneros bacterianos como *Kluyvera*, *Salmonella*, *Pseudomonas*, *Vibrio* o *Klebsiella*, pero la similitud se localizaba únicamente en el extremo amino terminal, en el dominio a través del cual se produce la unión de la fibra a la cola del fago. Este dominio pertenece a la familia PHA00430, siendo del que se pueden encontrar más secuencias similares, probablemente debido a que la mayor parte de los fagos de tipo T7 presentan mecanismos similares para unir la fibra al resto del fago. Por el contrario, el dominio punta es el más discrepante, obteniéndose menos secuencias similares. Se ha observado en las secuencias de varios fagos que, como respuesta adaptativa, se modifica el extremo de unión al receptor, permitiendo adaptarse a nuevos hospedadores y sortear los mecanismos defensivos que pudieran haber aparecido en sus anteriores hospedadores (Veesler y

Cambillau, 2011). Esta respuesta puede producirse mediante mutaciones aleatorias, pero también se han observado casos en los que se ha producido una transferencia horizontal del dominio de unión a receptor entre fagos, incrementando la variabilidad en esta región (Veesler y Cambillau, 2011).

Con los resultados obtenidos en el alineamiento de secuencias se puede observar que, en el caso de las fibras de los fagos que se utilizaron, las mutaciones son puntuales, seleccionadas en función de la mejora en la infectividad del fago frente a un hospedador concreto. Otras secuencias que se descartaron en el trabajo, como la de la espina del fago K1F, son un ejemplo del otro mecanismo adaptativo descrito, el intercambio de dominios, existiendo una alta identidad en el extremo amino terminal de unión de la fibra a la cola del fago, pero con extremos carboxilo terminales completamente diferentes, tanto que la punta de la espina del fago K1F está especializada en la degradación del polisacárido de las cápsulas de *E. coli* K1, presentando para ello actividad endo-N-acetilneuraminidasa (Schulz *et al.*, 2010b), una función completamente diferente a la de la fibra de T7, que es incapaz de unirse a una cepa encapsulada.

Teniendo en cuenta sólo el fragmento gp17(371-553) del que se ha resuelto la estructura, se puede representar la superficie de esta estructura en función del grado de conservación de los aminoácidos que la componen (Figura 4.13). El resultado muestra que las zonas más expuestas en el dominio de unión a receptor están menos conservadas, pero también que la zona en la que se unen las tres cadenas de la proteína muestra también un grado de conservación menor, causado especialmente por la Ala518 que en el resto de fagos es una serina o una leucina (Figura 4.13). Gran parte de este fragmento (aminoácidos 359-524) pertenece al dominio Pfam gp37_C, en el que se engloban gran parte de las regiones carboxilo terminales de fibras tipo gp37 de fagos tipo T4, pero curiosamente no incluye gp37 de T4. Esto podría indicar que estructuras similares al dominio pirámide de la fibra de T7 podrían formar parte de las fibras de otros podovirus y myovirus, ya que esta secuencia puede haber sido transferida de una familia a otra por transferencia horizontal y posteriormente se habría producido una divergencia de secuencia, pero siempre manteniendo un tipo de plegamiento que garantiza la función de la proteína.

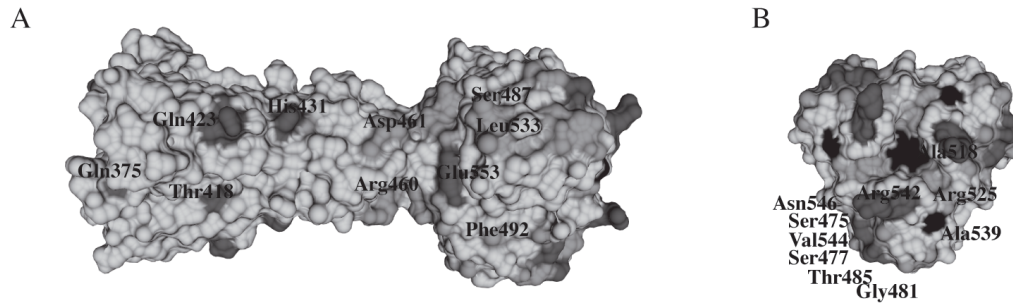


Figura 4.13: Representación en escala de grises de la conservación de los aminoácidos en la secuencia del fragmento gp17(371-553). Los residuos más oscuros son los menos conservados. (A) Vista lateral. (B) Vista frontal.

Para poder comparar la topología de la estructura con otras estructuras conocidas se realizó una búsqueda en el servidor DALI. No se obtuvieron resultados con la misma topología que estuvieran depositados en el PDB, pero sí existen otras proteínas que presentan similitud estructural. Al dominio punta se parece el extremo carboxilo terminal de la aguja del fago Sf6 (Bhardwaj *et al.*, 2011), y en menor medida a la fibra del bacteriófago PRD1 (Merckel *et al.*, 2005) y las fibras de adenovirus (van Raaij *et al.*, 1999) y reovirus (Guardado Calvo *et al.*, 2005) (Figura 4.14). Estas proteínas también están formadas por láminas β y su tamaño es similar al dominio punta de gp17. Analizando en detalle el plegamiento de la aguja del fago Sf6, si tenemos en cuenta la región comparable con gp17, existe una hélice α al igual que en gp17, seguida de cadenas β . La conexión entre algunas de las cadenas varía, estando formadas por las cadenas BIDG y CHEF en la aguja de Sf6, por lo que los lazos que conectan las cadenas en la parte superior son siempre entre cadenas que no están en la misma lámina, al contrario que en gp17, en la que sólo lo son la mitad (Figura 4.14A). Respecto a los elementos de la estructura, además de aquellos presentes en gp17, existen dos hélices α y dos cadenas β extra. Pese a la similitud en esta región, el resto de la aguja presenta una estructura diferente a la de gp17, en forma de triple hélice superenrollada, mientras que en las fibras de adenovirus, de reovirus y del bacteriófago PRD1, desde la región amino terminal al dominio cabeza el plegamiento es en triple β espiral. En cambio, las proteínas implicadas en la unión al receptor de los fagos p2 o TP901-1 de *Lactococcus* presentan dominios carboxilo terminales que también son similares, aunque en menor medida, a gp17, y su dominio asta es β -helicoidal (Spinelli *et al.*, 2006a; Spinelli *et al.*, 2006b).

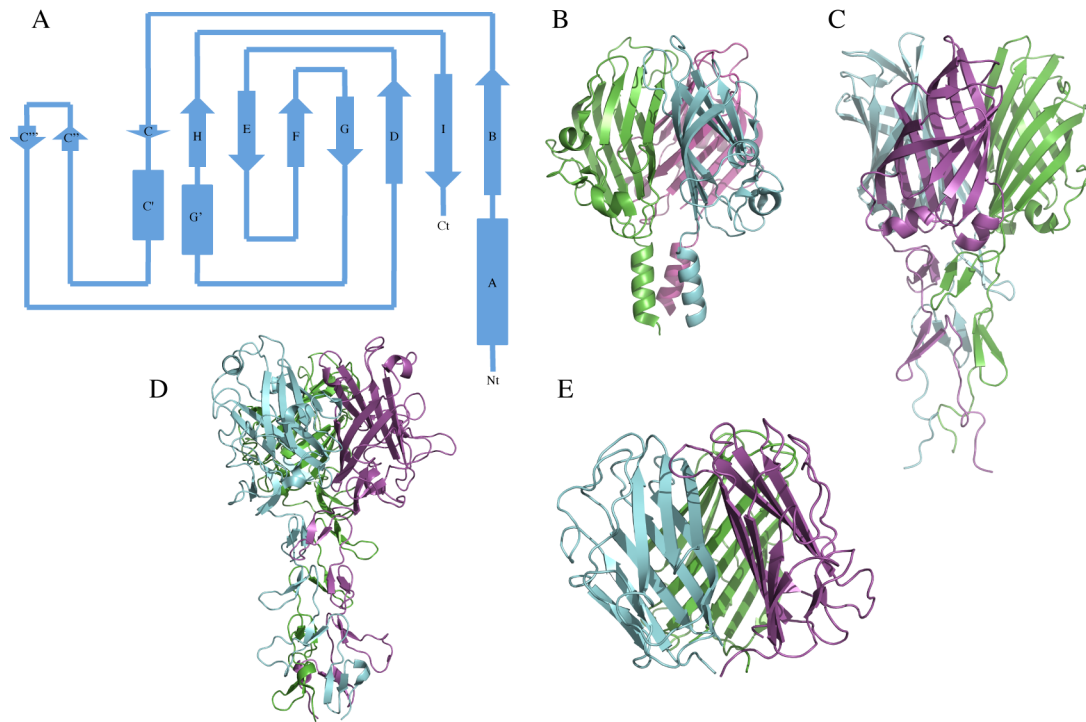


Figura 4.14: Comparación de la estructura del dominio punta de gp17(371-553) con otras estructuras. (A) Diagrama de topología del extremo carboxilo terminal de la aguja del fago Sf6. (B) Estructura del extremo carboxilo terminal de la aguja del fago Sf6. (C) Estructura del extremo carboxilo terminal de la fibra de reovirus aviar. (D) Estructura del extremo carboxilo terminal de la fibra de adenovirus. (E) Estructura del extremo carboxilo terminal de la fibra del fago PRD1. Las figuras B-E se corresponden con fragmentos carboxilo terminales.

Por otra parte, teniendo en cuenta que la morfología externa es similar, se puede comparar la estructura de gp17 con las fibras de otros fagos. Así pues, es destacable que tanto las fibras cortas como las fibras largas de T4 presentan plegamientos diferentes al de gp17. En el caso de las fibras largas, la proteína que forma su extremo carboxilo terminal es gp37, y el plegamiento que presenta consiste en cadenas β entrelazadas. En el caso de las fibras cortas, formadas por gp12, la región asta está formada por triple- β -hélices. Estas triple- β -hélices son uno de los plegamientos similares al del dominio pirámide, ya que tampoco se han localizado estructuras que presentasen exactamente la misma topología. El plegamiento en triple- β -hélice no sólo está presente en gp12, también se ha descrito en la aguja de T4 (gp5) (Kanamaru *et al.*, 2002), en la endosialidasa (Stummeyer *et al.*, 2005) y la endo-N-acetilneuraminidasa de K1F (Schulz *et al.*, 2010b), en la aguja de P22 (Olia

et al., 2007) y en la espina de $\phi 29$ (Xiang *et al.*, 2009). Pero sorprendentemente las proteínas con plegamientos más similares encontradas no son fibras ni agujas, son las puntas de los tubos internos de la cola de los myovirus P2 y $\phi 92$, gpV y gp138 respectivamente (Yamashita *et al.*, 2011; Browning *et al.*, 2012). Ambas estructuras se corresponden exclusivamente con el dominio pirámide, y las dos presentan la peculiaridad de tener un átomo de hierro en su extremo, careciendo de un dominio globular equivalente al dominio punta (Figura 4.15B y C). De estas dos proteínas, la que presenta una topología más parecida a gp17 es gp138, ya que independientemente del número de cadenas β , las conexiones entre cadenas del mismo monómero son siempre de forma antiparalela, y paralela entre cadenas de distintos monómeros (Figura 4.15A). Respecto al número de cadenas, empezando desde el carboxilo terminal, gp138 tiene 3 de un monómero y 7 de otro, mientras que en el caso de gp17 serían 3, 5 y 1. En cualquier caso, aunque el plegamiento no sea exactamente igual, resulta evidente que, como se ha discutido previamente, las proteínas que presentan plegamientos similares también tienen funciones similares en la célula, ya que todas las estructuras con las que se ha encontrado cierto grado de similitud son fibras, espinas, agujas o tubos internos, y todas ellas pertenecen a virus.

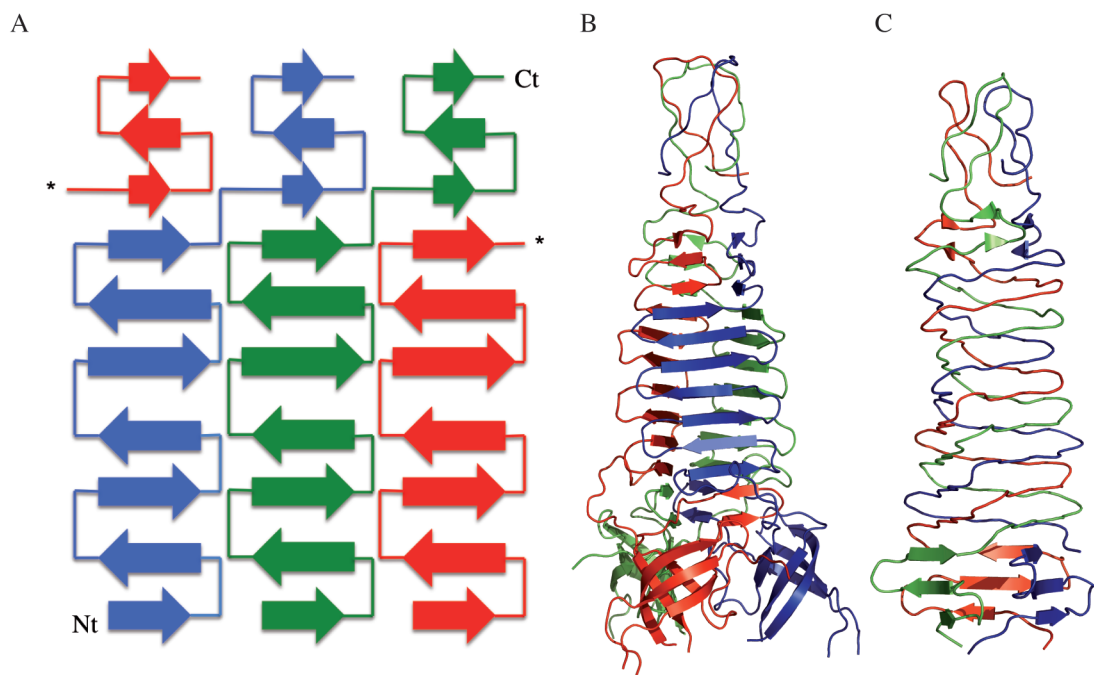


Figura 4.15: Comparación de la estructura del dominio pirámide de gp17(371-553) con otras estructuras. (A) Diagrama de topología de gp138. (B) Estructura de gp138. (C) Estructura de gpV.

4.4 La cola del bacteriófago T5

Las fibras del bacteriófago T5 no han sido tan estudiadas como las del fago T7. Al no disponer de imágenes de microscopía en las que se pudiera observar con claridad la fibra, se decidió llevar a cabo un estudio de las fibras mediante microscopía electrónica. Para poder afirmar que el plegamiento obtenido no es un artefacto de la sobreexpresión de la fibra en bacterias, se utilizaron colas completas de fagos. La cola de T5 está formada por varias proteínas, de las que destacan pb6 y pb2 por su alto número de copia. De la base de la cola parten las tres fibras, que serán el objeto final de estudio en esta sección.

4.4.1 Purificación de colas de T5

Para obtener las colas de T5 se creció el mutante *amber T5am20* en la cepa *E. coli* F tal y como se explicó en la sección 3.10.1. Tras los pasos iniciales de purificación, se realizó un gradiente de glicerol de doce fracciones en el que se identificaron aquellas en las que estaban presentes las colas (Figura 4.16A). Para ello, se utilizó como referencia la proteína más abundante de la cola, pb6, cuyo peso molecular es de aproximadamente 51 kDa. Se seleccionaron las fracciones de la 5 a la 11 para el siguiente paso de purificación, en una cromatografía de intercambio iónico. Se obtuvo un pico que eluyó a una concentración aproximada de 200 mM de cloruro sódico (Figura 4.16B) que se correspondía con las colas, que se identificaron de nuevo por la presencia de la proteína pb6, seleccionándose las fracciones de la 13 a la 18 (Figura 4.16C).

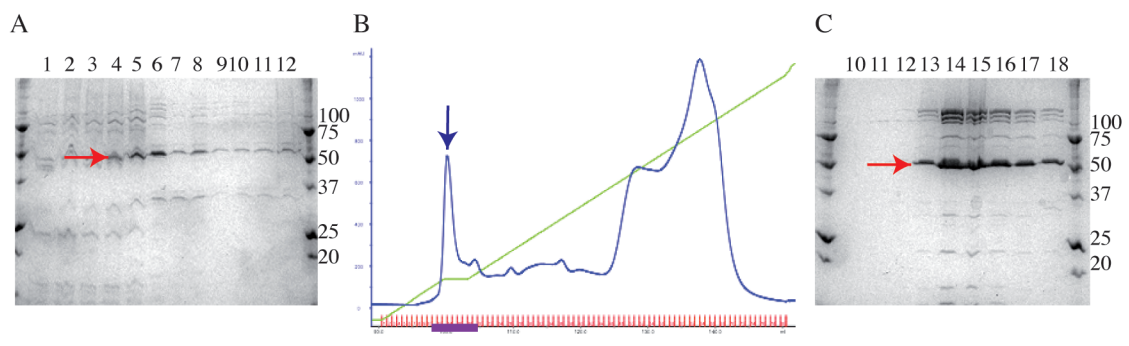


Figura 4.16: Proceso de purificación de las colas del fago T5. (A) Gradiente de glicerol. La flecha roja indica pb6. (B) Cromatografía de intercambio iónico. La flecha azul indica el pico en el que aparecen las colas. En morado se subrayan las fracciones que se seleccionaron para el gel. (C) Gel de las fracciones que contenían las colas en la cromatografía anterior. La flecha roja indica pb6.

4.4.2 Microscopía electrónica de las fibras de la cola de T5

Al disponer de un mutante sin cápside de T5, se pudieron obtener imágenes de microscopía electrónica en las que se observasen las fibras. Cuando se toman imágenes del virus completo, generalmente la tinción en la región de la cápside dificulta que exista contraste suficiente como para poder observar las fibras.

En las imágenes obtenidas mediante tinción negativa se puede diferenciar perfectamente la fibra, viéndose además que existe cierta rigidez en las partes proximal y distal de la fibra, pero una gran flexibilidad en la unión entre ambas regiones, que impide hacer un promediado de la fibra en conjunto. La parte proximal aparece como un asta muy fina, mientras que la parte distal muestra nódulos, de forma similar a los observados en las imágenes de microscopía electrónica de la fibra del fago T7. Además de las tres fibras, se puede diferenciar la espina central (Figura 4.17A).

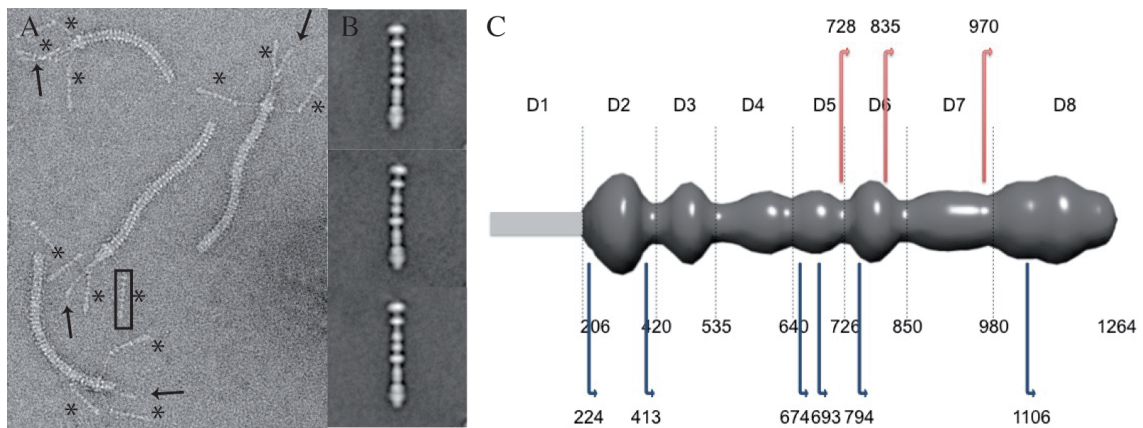


Figura 4.17: Microscopía electrónica de las fibras de la cola de T5. (A) Imagen tomada mediante tinción negativa de las colas: las flechas indican las espinas centrales y los asteriscos las fibras. El recuadro indica la región de las fibras que se seleccionó para ser promediada. (B) Tres diferentes promediados de la parte carboxilo terminal de la fibra. (C) Modelo obtenido por rotación. Se señalan los ocho dominios que se han descrito así como los fragmentos carboxilo terminales diseñados. En azul se señalan los que se han diseñado en función de la estructura secundaria y en rojo los diseñados basándose en el modelo obtenido por microscopía.

Una vez seleccionadas las imágenes correspondientes a la parte carboxilo terminal de la fibra, se realizaron promediados de la misma (Figura 4.17B) que permitieron construir un volumen por rotación. Este volumen se dividió en siete dominios

teniendo en cuenta las regiones en las que la fibra se estrecha más. Se calculó el volumen total y el de cada uno de los subdominios, pudiendo determinarse el comienzo aproximado de cada uno de ellos (Figura 4.17C) al asumir una distribución lineal de los aminoácidos en la fibra. Se asignó al primer dominio (D1) la parte proximal de la fibra, que abarca los aminoácidos 1-205 según la predicción de estructura secundaria descrita en la sección 4.5.1, y se asignaron los aminoácidos que forman cada uno de los siete dominios en los que se había dividido la parte distal: D2 (206-219), D3 (220-534), D4 (535-639), D5 (640-725), D6 (726-849), D7 (850-979) y D8 (980-1264).

Teniendo en cuenta el comienzo de los últimos tres dominios, se pudieron diseñar tres de los fragmentos carboxilo terminales, complementando el análisis con los resultados de predicción de estructura secundaria descritos más adelante (página 102).

4.5 La proteína *pb1* del bacteriófago T5

Pb1, codificada por el ORF124, conforma las fibras de la cola con forma de L. Este gen tiene 4191 pb, que codifican una proteína de 1396 aminoácidos. El fago tiene tres fibras y cada una de ellas está constituida por un trímero de pb1. Pb1 se procesa de forma que el dominio carboxilo terminal (CTD) de 132 aminoácidos se autoproteoliza tras ayudar al plegamiento de la proteína actuando como chaperona intramolecular, dando lugar a una fibra de 1264 aminoácidos. Existe cierta controversia acerca del punto exacto en el que se produce la proteólisis. Se ha descrito que ocurre entre la serina (1264 en el caso de T5) y el aspártico o asparagina que se encuentra a continuación (Schwarzer *et al.*, 2007; Xiang *et al.*, 2009), en otro trabajo más reciente se ha postulado que el corte se realiza antes de la serina, formando ésta parte del dominio chaperona (Schulz *et al.*, 2010a). En esta tesis se ha asumido que dicha serina podría formar parte de la fibra, por lo que se hace referencia a la fibra como pb1(1-1264).

4.5.1 Predicción de estructura secundaria

Los resultados obtenidos del servidor de predicción de estructura secundaria representados en la Figura 4.18 muestran que la fibra presenta tres regiones diferenciadas. La primera parte de la fibra comprende aproximadamente los 200 primeros aminoácidos, que la predicción indica que estaría formada por regiones

desestructuradas conectadas por pequeñas hélices α . Por otra parte, teniendo en cuenta que la proteína forma un trímero, es probable que estas regiones se plieguen en forma de hélice superenrollada o con plegamiento de colágeno, ya que son los plegamientos más típicos de esta clase de fibras. Esta región (1-205) se ha asociado con la parte proximal de la fibra. A partir de este punto la predicción cambia y la proteína se ordena principalmente en cadenas β casi hasta el extremo carboxilo terminal. A diferencia de la primera región, en este caso la estructura es mucho más compacta, siendo las regiones que conectan las cadenas β más cortas que en el caso anterior. Así pues, los aminoácidos 206-1264 se corresponderían con la parte distal de la fibra. El último fragmento vuelve a mostrar la presencia de hélices α , esta vez intercaladas con cadenas β . Esta última región mixta se corresponde con el dominio carboxilo terminal que actúa como chaperona intramolecular. La predicción es coherente con la estructura habitual de este tipo de fibras, con una parte amino terminal en la que predominan las hélices α y que es claramente elongada, frente a una parte carboxilo terminal formada principalmente por cadenas β mucho más compacta. Esta división en dos regiones es similar a la que presenta la fibra del bacteriófago T7.

La chaperona intramolecular presente en el extremo carboxilo terminal se autoproteoliza tras la serina 1264 al haberse plegado la fibra, aunque como se ha dicho previamente, este corte podría ocurrir antes de la serina. Este dominio había sido previamente descrito en la espina del fago K1F (Schulz *et al.*, 2010b) y los apéndices del fago $\phi 29$ (Xiang *et al.*, 2009), que presentan un dominio similar al de la estructura de la fibra de T5. El extremo carboxilo de la espina del fago K1F, también conocida como endosialidasa por su actividad catalítica, ha sido resuelto junto con el dominio chaperona (Schulz *et al.*, 2010b). Ya que las predicciones indican que en el caso de la fibra de T5 el dominio chaperona sería similar al de la espina del fago K1F debido a la gran similitud en su secuencia, se podría pensar que actuarían de forma similar. En la endosialidasa del fago K1F, el dominio chaperona se pliega rodeando la espina. El plegamiento está dominado por hélices α , a excepción de tres horquillas β , que son las que rodean el asta de la espina. Según se ha descrito, estas tres horquillas mantendrían las tres cadenas de la espina próximas y facilitarían el plegamiento en forma de triple- β -hélice. La predicción de este

Estructura de fibras de bacteriófagos

plegamiento es bastante consistente con la predicción de estructura secundaria obtenida para la región correspondiente de la fibra de T5.

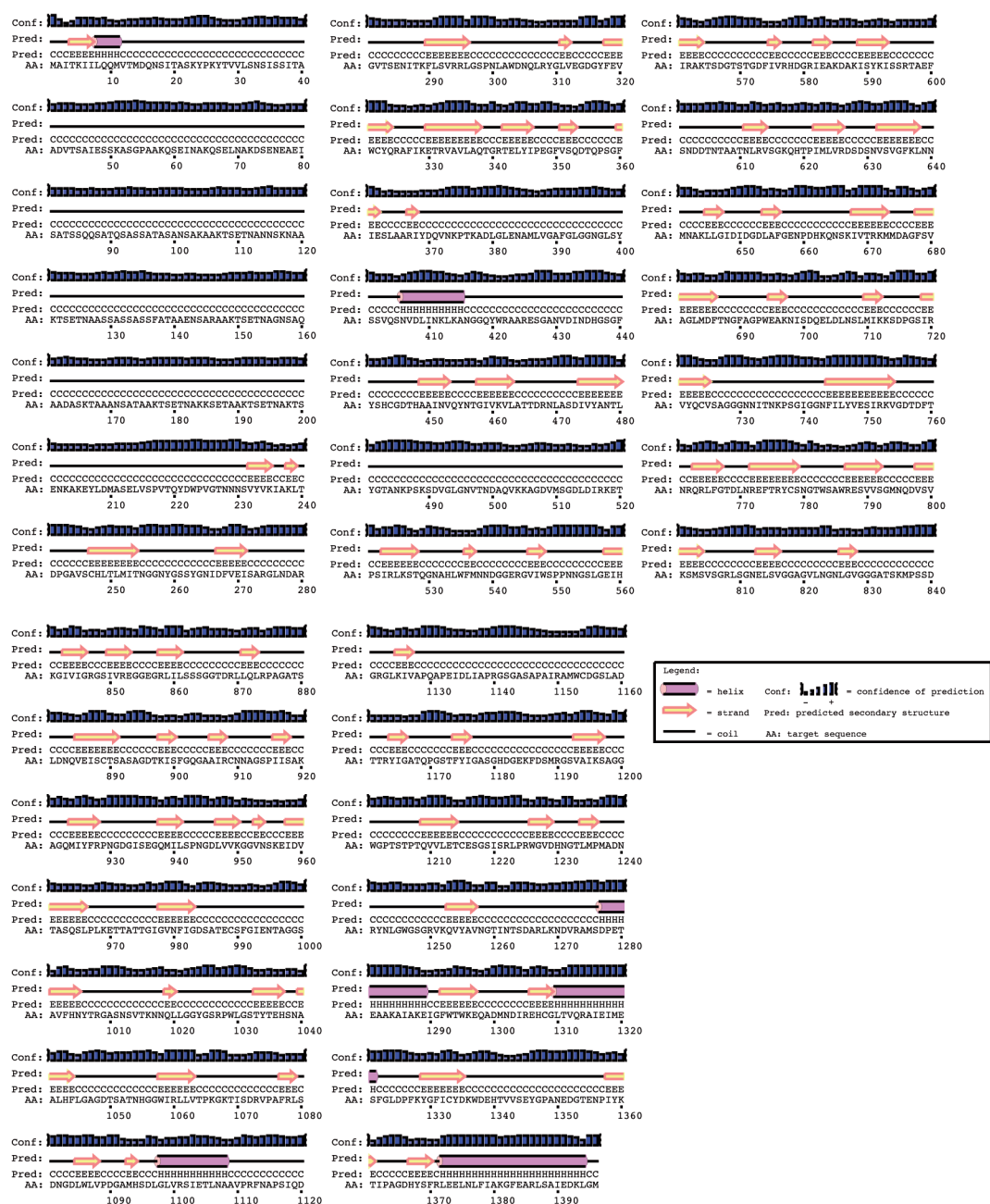


Figura 4.18: Predicción de estructura secundaria de pb1. Los aminoácidos se representan con códigos de una letra.

Los resultados obtenidos con la predicción de la estructura secundaria de pb1 permitieron diseñar diferentes fragmentos carboxilo terminales, complementados con los resultados diseñados en función de los resultados obtenidos por microscopía electrónica.

La división de la parte distal de la fibra en siete dominios en función de su volumen se ha comparado con los resultados obtenidos en la predicción de estructura secundaria. Las regiones en las que la fibra se estrecha conectando los dominios entre sí se corresponden en las predicciones con regiones desestructuradas de longitud mayor, por lo que el tipo de plegamiento sería diferente al de las regiones que conforman el interior de los dominios.

4.5.2 Clonaciones de fragmentos de pb1

Se clonaron los fragmentos de DNA correspondientes a los fragmentos carboxilo terminales de la proteína pb1 según se ha descrito en Materiales y Métodos, obteniendo los vectores pET28-pb1(674-1396), pET28-pb1(693-1396), pET28-pb1(794-1396), pET28-pb1(1106-1396), pET28-pb1(728-1396), pET28-pb1(835-1396) y pET28-pb1(970-1396) (Figura 4.19).

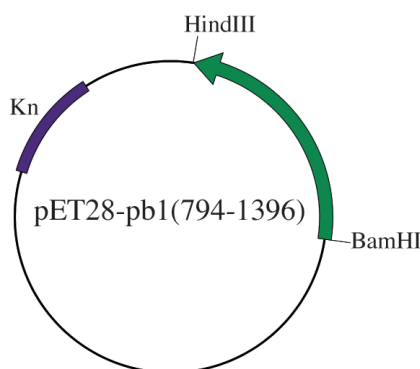


Figura 4.19: Modelo de vector obtenido para uno de los fragmentos de pb1.

4.5.3 Expresión de los fragmentos de pb1

Todos los fragmentos de pb1 clonados expresaban correctamente, aunque presentaban diferentes rendimientos y diferente grado de solubilidad (Figura 4.20). Con los fragmentos pb1(674-1396), pb1(693-1396), pb1(728-1396) y pb1(835-1396) se obtuvo expresión, pero en cantidades insuficientes para llevar a cabo ensayos de cristalización. El fragmento pb1(835-1396) se obtenía mayoritariamente en la fracción insoluble, por lo que es posible que el plegamiento no se estuviese llevando a cabo de forma correcta. Por otra parte, el fragmento pb1(728-1396) daba lugar a dos bandas, una de ellas en la que se había llevado a cabo la proteólisis del dominio chaperona intramolecular carboxilo terminal y otra en la que no, lo cual indicaba nuevamente problemas en el plegamiento. Con los fragmentos pb1(794-1396) y pb1(1106-1396) se obtuvo expresión soluble en cantidades suficientes y

correctamente procesado autoproteolíticamente, por lo cual se llevaron a cabo purificaciones y ensayos de cristalización. En cualquier caso, los mejores resultados se obtuvieron con el fragmento pb1(970-1396) se obtuvo en cantidades suficientes como para permitir realizar los ensayos de cristalización (Figura 4.20). Los resultados obtenidos con este último fragmento concuerdan con las predicciones realizadas mediante microscopía electrónica y la predicción secundaria, ya que este fragmento abarcaría el último dominio y la chaperona (que se autoproteolizaría) y, por lo tanto, sería el más compacto y globular, en el que en la predicción secundaria no aparecen grandes regiones que pudiesen formar lazos, que podrían provocar problemas tanto de agregación durante la purificación como a la hora de la cristalización.

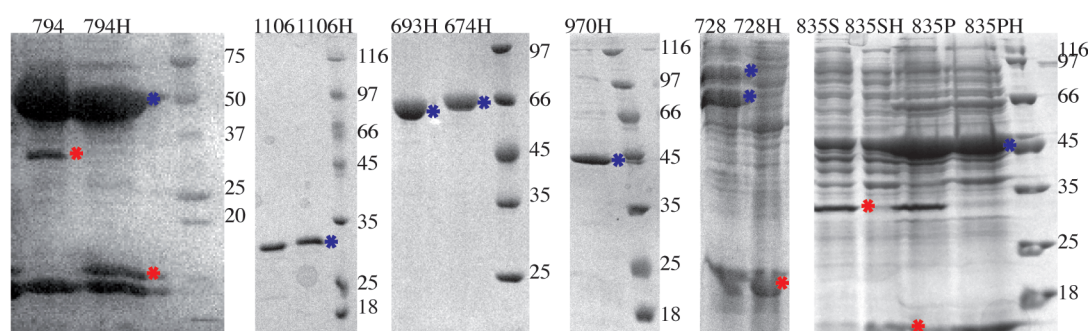


Figura 4.20: Muestras de expresión y/o purificación de cada uno de los fragmentos de pb1 clonados. En cada gel se indica el aminoácido en el que comienza el fragmento así como si la muestra ha sido hervida (H) o no. En el último caso, el gel incluye muestras del sobrenadante (S) y el pellet (P). Los asteriscos azules indican las bandas correspondientes a los fragmentos de la fibra; los rojos, el fragmento carboxilo terminal proteolizado.

4.5.4 Purificaciones de fragmentos de pb1

Se llevaron a cabo purificaciones de los fragmentos pb1(794-1396), pb1(1106-1396) y pb1(970-1396). En los tres casos se observó que el dominio carboxilo terminal se autoproteolizaba como estaba previsto, generando por lo tanto dos bandas en los geles, una correspondiente a la parte de la fibra y un fragmento pb1(1265-1396) (o pb1(1264-1396), dependiendo del punto exacto de corte) correspondiente a la chaperona intramolecular (15 kDa). El proceso de purificación realizado fue equivalente para los tres fragmentos. Para seleccionar las fracciones con las que se continuaría el proceso de purificación, se tuvo en cuenta la presencia de la fibra,

independientemente de que en algunos casos copurificase con el dominio carboxilo terminal.

En el caso del fragmento pb1(794-1264) se realizó una cromatografía de afinidad, en la que la proteína eluyó entre 50 y 500 mM de imidazol (Figura 4.21A). Las fracciones recogidas se utilizaron para un segundo paso de cromatografía de intercambio iónico, obteniéndose un pico a una concentración de 125 mM de cloruro sódico (Figura 4.21B y Figura 4.21C). En todas las purificaciones se observaba una banda débil que se correspondía con pb1(1265-1396), lo que podría indicar que parte de las fibras todavía están interaccionando con este dominio, aunque la proteólisis se ha producido ya que se obtienen dos bandas independientes. Pese a que este factor introduce variabilidad en la muestra, se utilizó la proteína purificada para realizar ensayos de cristalización.

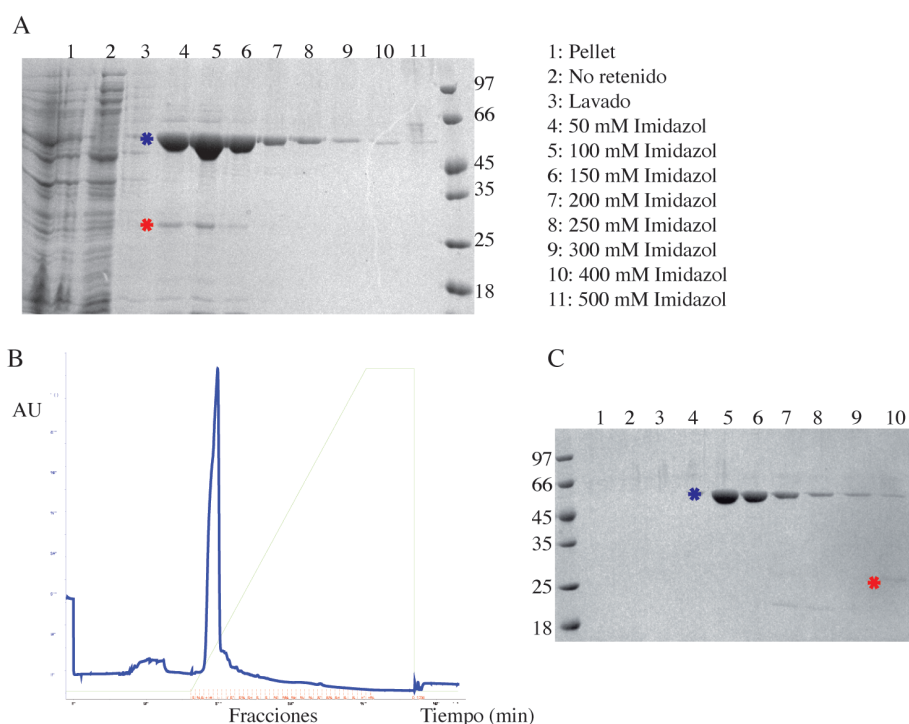


Figura 4.21: Purificación del fragmento pb1(794-1264). (A) Gel con las fracciones obtenidas de la cromatografía de afinidad. (B) Cromatograma de interacción iónica. (C) Gel con las fracciones recogidas en la cromatografía anterior. Los asteriscos azules indican las bandas correspondientes a los fragmentos de la fibra; los rojos, el fragmento carboxilo terminal proteolizado.

El comportamiento del fragmento pb1(1106-1264) era similar al de pb1(794-1264). En la cromatografía de afinidad la elución se produce en el mismo rango 50-

Estructura de fibras de bacteriófagos

500 mM de imidazol (Figura 4.22A), pero en el caso de la cromatografía de intercambio iónico la elución se retrasa hasta una concentración de 280 mM de cloruro sódico (Figura 4.22B). El pico obtenido no fue tan claro y diferenciado como en el caso del fragmento pb1(794-1264) pero en este caso, al analizar en gel, se observa que la banda correspondiente a pb1(1265-1396) está presente en más fracciones, por lo que es probable que la muestra sea más homogénea (Figura 4.22C), puesto que se observan cantidades similares de chaperona y de fibra. La proteína purificada se utilizó para ensayos de cristalización, sin obtenerse cristales.

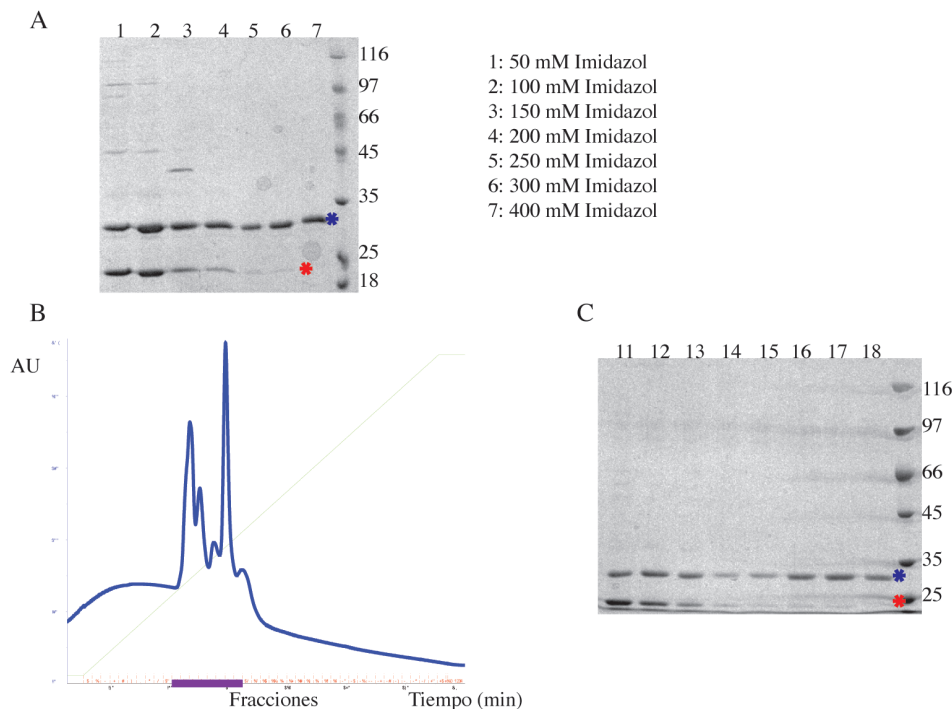


Figura 4.22: Purificación del fragmento pb1(1106-1264). (A) Gel con las fracciones obtenidas de la cromatografía de afinidad. (B) Cromatograma de interacción iónica. Se señalan en violeta las fracciones seleccionadas. (C) Gel con las fracciones recogidas en la cromatografía anterior. Los asteriscos azules indican las bandas correspondientes a los fragmentos de la fibra; los rojos, el fragmento carboxilo terminal proteolizado.

Por último, el fragmento pb1(970-1264) muestra un comportamiento ligeramente diferente. La fibra queda retenida en la cromatografía de afinidad durante más tiempo, empezando a eluir a partir de 150 mM de imidazol. En el caso de la cromatografía de intercambio iónico, se obtiene un pico múltiple a una concentración de 300 mM de cloruro sódico. La proteína no presenta un grado de pureza tan elevado como en casos anteriores, pero la cantidad de proteína obtenida es

aproximadamente tres veces mayor. Se sigue observando la presencia de la chaperona copurificada con la fibra y, además, se observa que en este caso la fibra mantiene parcialmente su condición trimérica en el gel cuando las muestras no se hierven.

El fragmento pb1(970-1264) ha sido también expresado en la cepa B834(DE3) para obtener proteína con selenio en lugar de azufre en las metioninas. Los resultados obtenidos durante la purificación utilizando esta cepa auxótrofa son equivalentes a los previos utilizando BL21(DE3), aunque se obtuvieron mayores cantidades de proteína. En la columna de afinidad se produce la elución a partir de 150 mM de imidazol (Figura 4.23A), y en la cromatografía de intercambio iónico la elución ocurre a una concentración mayor de cloruro sódico comparado con la proteína nativa, aproximadamente a 500 mM (Figura 4.23B y Figura 4.23C), obteniéndose igualmente un pico múltiple. Como se obtuvo mayor cantidad de proteína en las purificaciones con la cepa auxótrofa, la mayor parte de ensayos de cristalización de este fragmento se han realizado directamente con la proteína modificada con selenio-metionina.

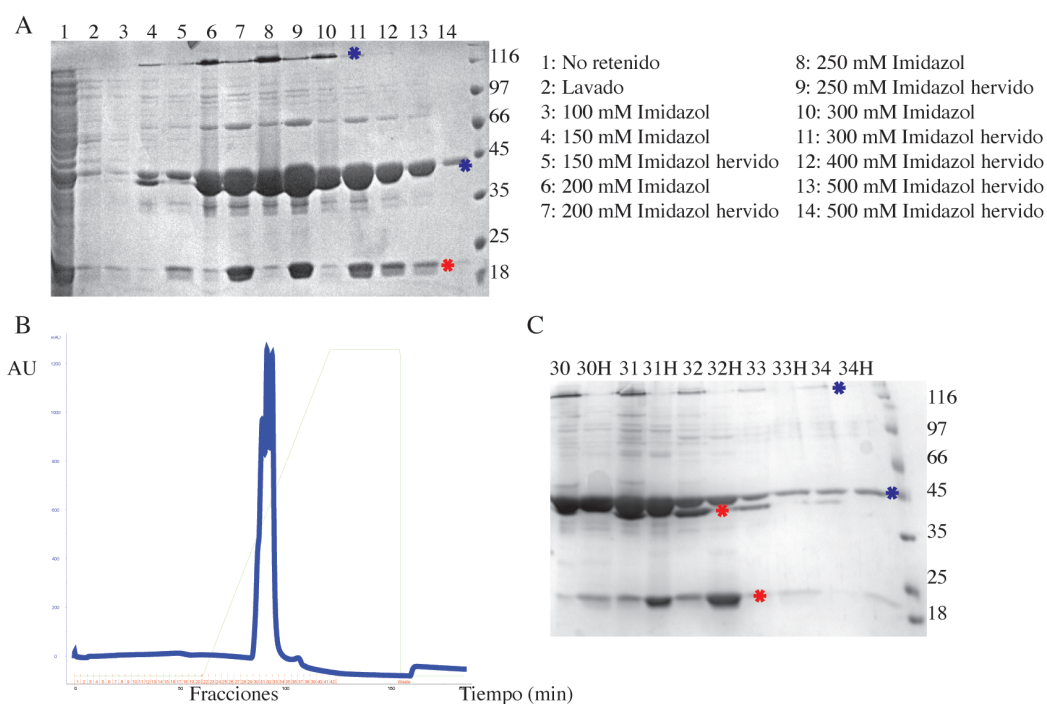


Figura 4.23: Purificación del fragmento pb1(970-1264). (A) Gel con las fracciones obtenidas de la cromatografía de afinidad. (B) Cromatograma de interacción iónica. (C) Gel con las fracciones recogidas en la cromatografía anterior, no hervidas y hervidas (H). Los asteriscos azules indican las bandas correspondientes a los fragmentos de la fibra; los rojos, el fragmento carboxilo terminal proteolizado.

4.5.5 Cristalización del fragmento pb1(794-1264)

Se obtuvieron cristales del fragmento pb1 (794-1264) en gotas cuya condición de cristalización era 0,1 M MES pH 6,5 y 2 M de sulfato amónico. Se probaron diferentes condiciones variando el tampón y el pH, variando la concentración de sulfato amónico y probando diferentes aditivos (Figura 4.24) hasta obtener los mejores resultados con cristales que difractaban a aproximadamente 2 Å (Figura 4.26A). Los cristales pertenecen al grupo espacial $P6_3$ y tienen una celdilla unidad de 52 x 52 x 298 Å. Aunque la resolución era buena, los patrones de difracción de estos cristales presentaban problemas a la hora de la integración, debido a la longitud del eje mayor (298 Å). Este problema implica que, dependiendo de la posición del cristal, se obtienen reflexiones que solapan entre ellas, o reflexiones que están tan próximas unas a otras que en lugar de un punto se observa una línea. En ambos casos es complicado determinar las intensidades de las reflexiones individuales, disminuyendo el número de datos que se pueden utilizar y dificultando el proceso de resolución de la estructura.

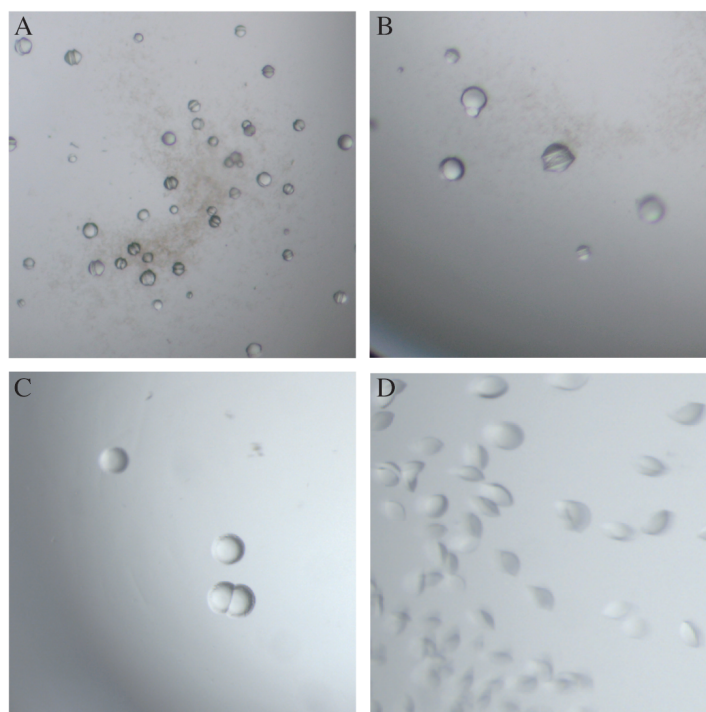


Figura 4.24: Diferentes morfologías obtenidas en los cristales del fragmento pb1(794-1264). Para cada uno de ellos varían las condiciones de cristalización: (A) 0,1 M HEPES pH 7,5, 1,5 M sulfato amónico. (B) 0,1 M MES pH 6,5, 1,5 M sulfato amónico. (C) 0,1 M MES pH 6,5, 1,5 M sulfato amónico, 2M tartrato sódico. (D) 0,1 M MES pH 6,5, 1,5 M sulfato amónico, 1M cloruro césico.

4.5.6 Cristalización del fragmento pb1(970-1264)

Basándose en los resultados obtenidos con el fragmento pb1(794-1264), se cristalizó este fragmento en gotas cuya condición de cristalización era 0,1 M Tris pH 8,5, 2 M sulfato amónico (Figura 4.25A), que difractaron a más de 3 Å (Figura 4.26B), con grupo espacial *R*32, una celdilla de 80 x 80 x 80 y un ángulo de 53°.

Como también se obtuvo proteína con selenio-metionina, se utilizaron las mismas condiciones de cristalización que para la proteína nativa, obteniendo cristales en varias gotas en las que el sulfato amónico había sido sustituido por PEG (Figura 4.25B), difractando el mejor cristal a 2,5 Å (Figura 4.26C). Estos cristales pertenecían al grupo espacial *C*2 y sus dimensiones de celdilla 227 x 58 x 70 con un ángulo β de 98°. Las estadísticas obtenidas se encuentran en la Tabla 4.7. Las predicciones indicaban la presencia de un trímero en la unidad asimétrica, con un contenido del 46% de solvente.

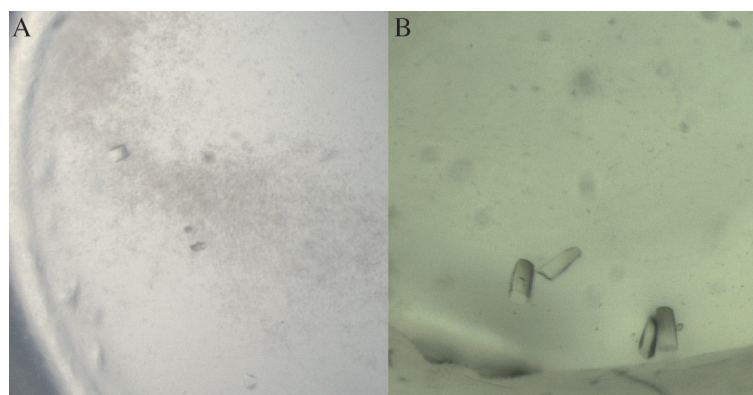


Figura 4.25: Cristales obtenidos del fragmento pb1(970-1264). (A) Proteína nativa. (B) Derivado con selenio.

Tabla 4.7: Estadísticas obtenidas del escalado del sets de datos utilizado para el SAD de pb1(970-1264).

Longitud de onda (en Å)	0,97973
Parámetros de la celdilla (a x b x c, Å)	227, 58 x 70
Rango de resolución (Å)	56,2-2,5 (2,59-2,52)
Reflexiones	30585 (2237)
Multiplicidad	6,3
Complejidad	96,8 (95,8)
$\langle I/\sigma(I) \rangle$	11,4 (2,3)
R_{merge} (%)	10,0 (64,4)

Los valores entre paréntesis se corresponden con la capa de máxima resolución.

$$R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$$

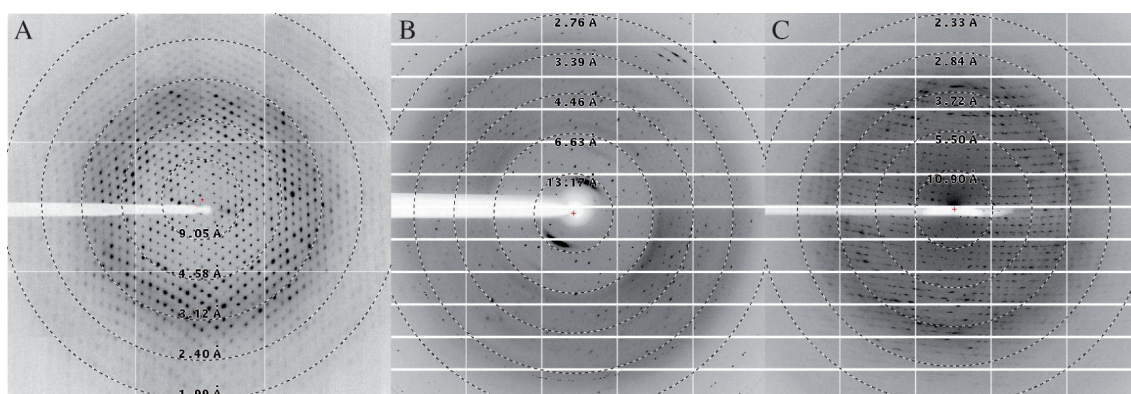


Figura 4.26: Patrones de difracción obtenidos para los dos fragmentos de pb1. (A) pb1(794-1264). (B) pb1(970-1264) nativo. (C) pb1(970-1264) con selenio.

Tabla 4.8: Estadísticas del faseado de pb1(970-1264).

Longitud de onda (en Å)	0,97973
Reflexiones (acéntricas)	28621
Átomos pesados	15 Se
Coefficiente de correlación (todos/débiles)	41,20/24,01
Figura de mérito de Patterson	12,07
Coefficiente de correlación (E)	0,294
R-Cullis [§]	
Anómalo (acéntrico)	0,780
Poder de faseado [§]	
Anómalo (acéntrico)	1,160
Figura de mérito (acéntrica/céntrica) [§]	0,327/0,128
Aplanamiento de solvente (28 ciclos con 46,4% de solvente) [¶]	
Factor R (antes/después de la modificación de densidad)	0,5000/0,1820
Correlación media de $ E ^2$ (antes/después de la modificación de densidad) [¶]	0,2850/0,6945
Puntuación de la mano = correlación de $ E ^2$ /contraste (original/inversa)	0,3176/0,4066

Encontrados con SHELXD (Sheldrick, 2008). [§]Calculado con SHARP (Vonrhein *et al.*, 2007). [¶]Según SOLOMON (Abrahams y Leslie, 1996).

Mediante la técnica de dispersión anómala simple (SAD) se localizaron las posiciones de los quince selenios localizados en las selenio-metioninas de la proteína, cinco pertenecientes a cada monómero. Las estadísticas del proceso se encuentran en la Tabla 4.8. El faseado y aplanamiento de solvente se hizo con los datos obtenidos a 2,5 Å seleccionando las fases de la mano inversa, pudiendo

construir automáticamente un modelo inicial parcial con 618 aminoácidos (206 en cada monómero). Este modelo se completó manualmente y se realizó un ajuste y refinado, obteniendo finalmente un modelo con 260, 271 y 267 aminoácidos en los monómeros A, B y C y 113 moléculas de agua observadas. No se ha observado densidad correspondiente a los aminoácidos 970-1003, 993-1003 y 997-1003 en los monómeros A, B y C respectivamente, ni para el aminoácido 1264 en ninguno de los tres monómeros. Así pues, teniendo en cuenta la simetría no cristalográfica, se ha denominado el fragmento presente en el modelo como pb1(993-1263). Las estadísticas del refinamiento se muestran en Tabla 4.9.

Tabla 4.9: Estadísticas del refinamiento de pb1(993-1263).

Número de reflexiones usadas	28484
Número de reflexiones usadas para el R_{free}	1897(200)
Factor R^{\S}	0,23 (0,27)
R_{free}	0,28(0,31)
Valor B medio/ Valor B Wilson (\AA^2)	49,5/41,1
Estadísticas del Ramachandran [¶] (%)	93,2/99,6
Rmsd ^{**} (enlaces, \AA /ángulos, °)	0,012/1,6

Los valores entre paréntesis se refieren a la capa de máxima resolución cuando es aplicable.

$^{\S}R = \sum ||F_{\text{obs}}(hkl)| - |F_{\text{calc}}(hkl)|| / \sum |F_{\text{obs}}(hkl)|$. $^{\¶}$ Determinado con MOLPROBITY (Chen *et al.*, 2010). Los porcentajes indican los residuos en regiones favorecidas y permitidas respectivamente. ** Obtenido con REFMAC (Murshudov *et al.*, 2011).

4.5.7 Descripción de la estructura

En la estructura se pueden diferenciar cuatro regiones, que se han nombrado de forma similar a las regiones de la espina de la cola del fago P22 debido a su similitud, que se explicará más adelante (página 121). La parte amino terminal está formada por un dominio triangular en forma de β prisma, formado principalmente por cadenas β antiparalelas que forman una lámina β cóncava con dos pequeñas hélices α intercaladas (aminoácidos 1010-1129). Después se encuentra una región entrelazada en la que una horquilla β de cada monómero se entrecruza con los monómeros adyacentes (aminoácidos 1130-1160). A continuación, está presente un

dominio triangular en forma de β prisma (aminoácidos 1161-1238), que lo componen cinco cadenas β antiparalelas que forman otra lámina β cóncava. Por último, el dominio aleta caudal (aminoácidos 1239-1263) está formado por cadenas β paralelas que se entrecruzan formando una triple- β -hélice (Figura 4.27A).

Teniendo en cuenta la presencia de amplios lazos conectando las cadenas β en la estructura, se calculó la diferencia (*rmsd*) entre los tres monómeros, siendo de 0,28 Å para los carbonos α y de 0,31 Å si se tienen en cuenta las cadenas laterales, encontrándose la mayor parte de las diferencias en los aminoácidos que se encuentran en el extremo amino terminal y no en los lazos que conectan las cadenas β . Para nombrar las estructuras presentes en la proteína y analizar su topología, se han denominado las dos pequeñas hélices α como $\alpha 1$ y $\alpha 2$ y las cadenas β alfabéticamente desde el extremo amino terminal, de forma que la última cadena β es la cadena Z. Así pues, el dominio triangular amino terminal estaría formado por las cadenas K-O, las dos hélices α y la cadena P. La región entrelazada por las cadenas Q y R, el segundo dominio triangular por las cadenas S-W y la aleta caudal por las cadenas Y-Z (Figura 4.27B).

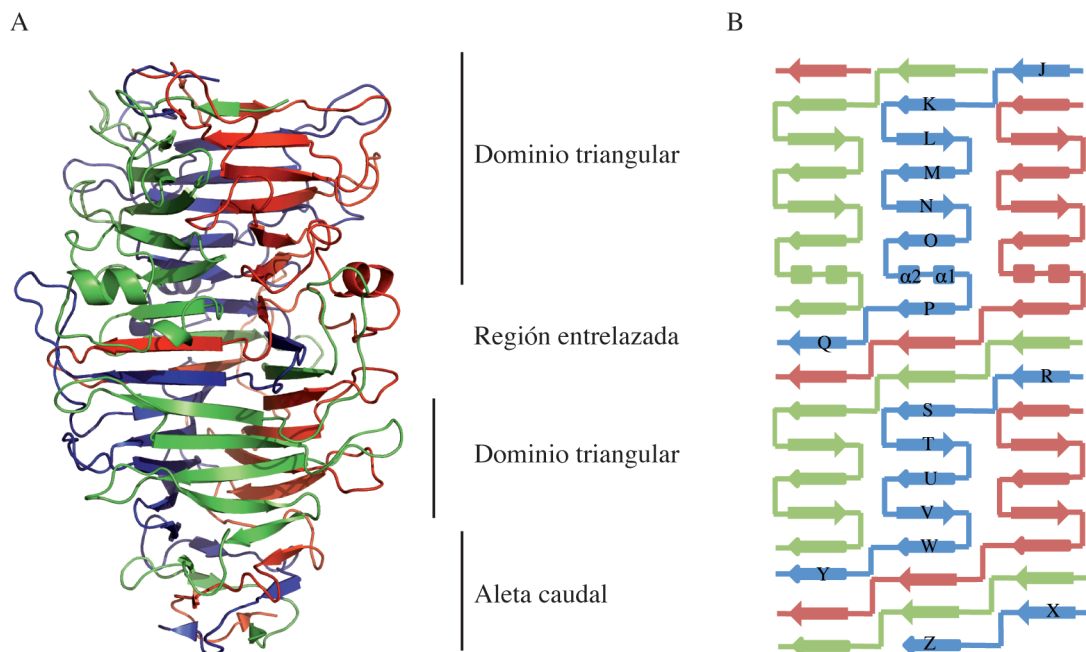


Figura 4.27: Estructura de pb1(993-1263). (A) Vista lateral de la estructura obtenida de pb1 con los diferentes dominios señalados. (B) Diagrama de la topología de la proteína, con las diferentes cadenas nombradas en uno de los tres monómeros.

El primer dominio triangular está formado por seis cadenas β . Las cadenas son antiparalelas entre sí, a excepción de las dos últimas, O-P, que son paralelas y están separadas por un largo lazo en el que se encuentran dos pequeñas hélices α , con dos y seis aminoácidos respectivamente. El conjunto de todas las cadenas forma una lámina β cóncava en la que se puede observar una tendencia a disminuir la longitud de las cadenas β a medida que se avanza hacia el extremo carboxilo terminal. En los lazos que conectan las cadenas β se observa la misma tendencia, a excepción obviamente de la conexión O-P, ya que en ella se encuentran las dos hélices α .

A continuación se encuentra la región entrelazada, en la que se forma una horquilla β con las cadenas Q-R. A modo de ejemplo, la cadena Q del monómero A se encuentra bajo la cadena P del monómero B y la cadena R se encuentra sobre la cadena S del monómero C. La presencia de dominios entrelazados como este se podría interpretar como un mecanismo para dotar de estabilidad a la fibra, ya que ayuda a mantener el trímero unido. También hay que destacar que pese a ello, el lazo de mayor tamaño de la estructura se encuentra justamente conectando las cadenas R y S, con una longitud de 19 aminoácidos, que a su vez podría proporcionar más movilidad a esta región.

El segundo dominio triangular se encuentra formando otro β prisma. Está formado por cinco cadenas β de cada uno de los monómeros, en disposición antiparalela, que forman tres láminas (una de cada monómero) con forma cóncava de forma que, junto con el otro dominio triangular, dota de aspecto triangular a la estructura. Las cadenas que forman este dominio (S-W) tienen una longitud decreciente hacia el extremo carboxilo, estando formadas las dos primeras formadas por siete aminoácidos, la tercera y la cuarta por seis, y la última por cuatro. La longitud de los lazos que conectan estas cadenas es variable, pero también con una tendencia decreciente.

Por último, en el dominio aleta caudal se encuentran entrecruzadas tres cadenas β de cada uno de los monómeros dando lugar a una triple- β -hélice y formando una lámina β en la que las tres cadenas son de pequeño tamaño.

En la descripción anterior no se ha incluido la parte amino terminal en la que se encuentra la cadena J. Esta cadena está entrecruzada con el siguiente monómero en su extremo amino, y continúa sobre el tercer monómero, por lo que indica que antes de la cadena J se encontraría una cadena I y ambas formarían otra región entrelazada. En cualquier caso, al no observarse claramente densidad para los aminoácidos que

formarían la cadena I, no se puede afirmar con rotundidad que se forme esta región entrelazada, por lo que se ha excluido de la descripción topológica de la estructura.

Analizando la composición aminoacídica de la proteína se observa que la mayor parte de los residuos hidrofóbicos se encuentran hacia el interior, creando un núcleo claramente hidrofóbico, y encontrándose los residuos hidrofílicos hacia el exterior y en los lazos. La mayor parte de los residuos cargados se encuentran en los lazos que conectan las cadenas, existiendo más cargas positivas (11%) que negativas (7.5%) (Figura 4.28), destacando especialmente las cargas negativas concentradas en la región entrelazada que forma la horquilla β y el lazo OP en el que están presentes las hélices α . Por otra parte, en la zona central de las cadenas J, L, M, N y P se localizan aminoácidos con carga positiva y, como estas cadenas forman una lámina cóncava, se genera un surco con carga positiva. En el dominio triangular formado por las cadenas S-W también se observa la formación de un surco, con cargas localizadas hacia los laterales, alternándose regiones con cargas negativas a los extremos del dominio con cargas positivas en la región central (Figura 4.28B).

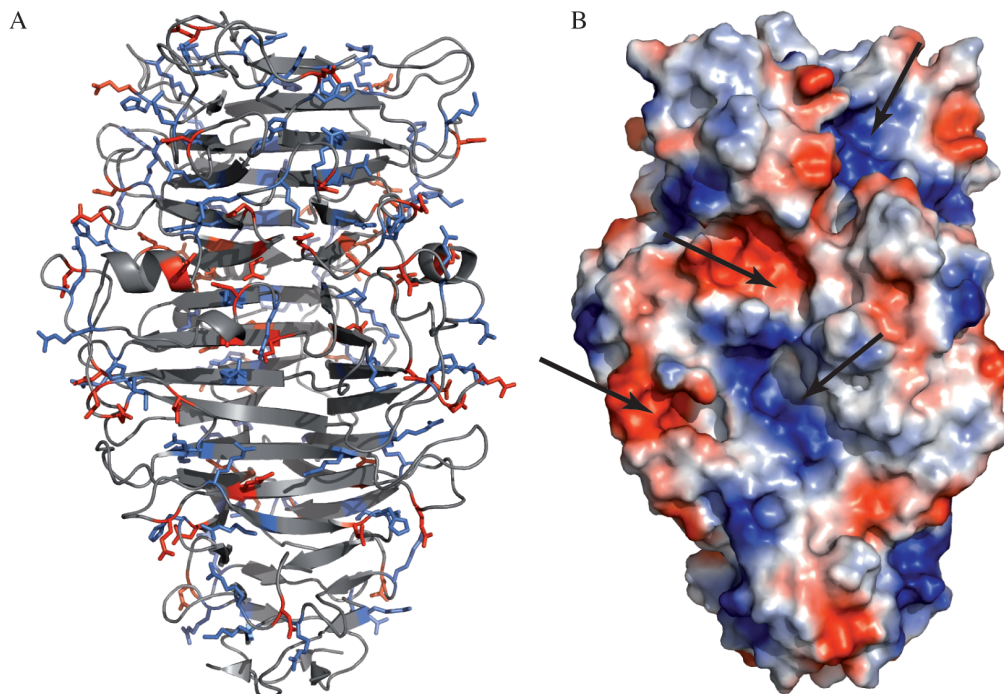


Figura 4.28: Representación de las diferentes cargas presentes en pb1. Los cargados positivamente aparecen representados en azul y los negativos en rojo. (A) Representación de las cadenas laterales. (B) Representación de la superficie electrostática. Se indican con flechas las regiones en las que se ha localizado una mayor concentración de aminoácidos cargados positiva o negativamente.

4.5.8 Estabilidad y plegamiento

Pese a la longitud de algunos de los lazos que conectan las cadenas β , se ha observado que su densidad es muy clara, por lo que se encuentran estabilizados. La única región en la que la densidad no es clara y donde ha sido complicado asignar la posición de las cadenas laterales ha sido en el extremo amino. Observando las moléculas relacionadas por simetría se puede asegurar que la mayor parte de aminoácidos que no se han asignado (aproximadamente 20 de la proteína más la cola de histidinas) no están presentes en la estructura, es posible que no se encuentren en la proteína ya que los últimos aminoácidos asignados se encuentran a corta distancia de los relacionados por simetría, aunque también podrían estar desordenados o haber sido proteolizados. En cualquier caso, una pequeña fracción de aminoácidos presentes en el cristal no han podido ser situados en la estructura debido al desorden en este extremo que impedía observar una densidad clara.

Se ha calculado la superficie expuesta en el trímero y para cada uno de los monómeros, pudiendo determinar que la superficie expuesta en el trímero es de $24 \times 10^3 \text{ \AA}^2$ y, analizando el monómero por separado, la superficie expuesta es de $19 \times 10^3 \text{ \AA}^2$. El 58% de la superficie queda enterrada al formarse el trímero, generando una superficie enterrada total de $33,4 \times 10^3 \text{ \AA}^2$. También se ha calculado el número de posibles enlaces, obteniendo 186 enlaces de hidrógeno y 35 puentes salinos intramonómero y 270 enlaces de hidrógeno y 50 puentes salinos intermonómero en cada monómero. Los enlaces intermonómero dotan de gran estabilidad a la proteína y dificultan su disociación, lo que se corroboró calculando la energía de disociación del trímero, que es de 297,5 Kcal/mol. Respecto a la posibilidad de formación de puentes disulfuro, no se han detectado en el cristal y, aunque existen dos cisteínas en cada monómero, la distancia entre ellas hace poco probable la formación de dichos puentes.

Por otra parte hay que destacar que el empaquetamiento de las moléculas en el cristal es muy compacto y genera contactos cristalinos. De esta forma, se han detectado distancias que favorecerían la formación de enlaces con otros trímeros presentes en el cristal, pudiendo formarse dos puentes salinos y siete enlaces de hidrógeno entre un trímero y los trímeros adyacentes.

En el dominio aleta caudal, las tres cadenas se encuentran entrecruzadas, pudiendo ser éste un punto de partida para la formación del trímero de la proteína.

Este modelo de trimerización se ha propuesto previamente en el caso de la espina de la cola del fago P22 (Kreisberg *et al.*, 2002). Al contrario que en P22, en T5 existe una chaperona intramolecular en el extremo carboxilo terminal que ayuda al correcto plegamiento y trimerización de la fibra. La secuencia de la chaperona de T5 es homóloga a la del fago K1F, por lo que es esperable que presenten un plegamiento similar. La estructura de la chaperona del fago K1F es conocida, caracterizándose por la presencia de un dominio de hélices α y otro dominio formado por tres horquillas β , que formarían una cavidad en la que se plegaría la fibra y a su vez abrazarían la fibra para facilitar el plegamiento. Una vez plegada la fibra en su extremo, se produciría una autoproteólisis tras la Ser1264 (en T5) que liberaría la chaperona. En la estructura aquí presentada no se ha observado densidad para la Ser1264, al igual que en otros trabajos en los que se ha planteado que la autoproteólisis ocurre antes de esta serina.

Previamente también se ha destacado la presencia de un surco con aminoácidos cargados en el primer dominio triangular. Si se observa la distribución de cargas en las horquillas de la chaperona del fago K1F, se pueden localizar cargas negativas que quedarían expuestas hacia el interior (hacia la fibra) correspondiéndose con la región en la que existen cargas positivas en el surco de la fibra de T5, por lo que es posible que ese surco positivo sirva como anclaje para las horquillas de la chaperona, que se unen a la fibra que se está formando para facilitar el plegamiento. Dada la elevada homología de secuencia entre la chaperona del fago K1F y la de T5, se ha realizado un *docking* entre dicha chaperona y la estructura de la fibra de T5. Como se puede observar en la Figura 4.29A, la fibra encaja en el hueco formado por la chaperona, y las horquillas β de la chaperona se sitúan sobre los surcos presentes en la fibra. Si se analiza el extremo carboxilo terminal de la fibra, en el que no se ha podido observar la Ser1264, se observa que la Thr1263 aparece orientada hacia el Asp912 de la chaperona del fago K1F, que ocupa la posición que correspondería al aminoácido 1265 en la chaperona de la fibra de T5, estando ambos aminoácidos separados por un espacio equivalente al del aminoácido que no está presente en la estructura de la fibra (Figura 4.29B). Dada la controversia acerca del punto exacto de proteólisis, aunque la Ser1264 no está presente en la fibra, tampoco se ha incluido en la chaperona del fago K1F.

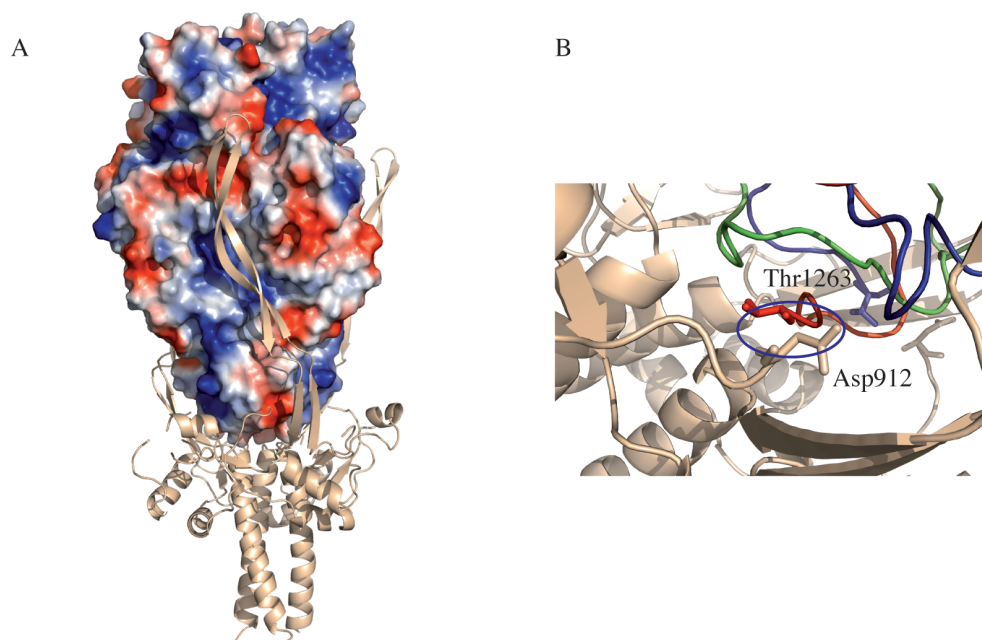


Figura 4.29: Interacción entre la proteína pb1 y la chaperona carboxilo terminal. (A) Representación de pb1 del fago T5 unida a la chaperona del fago K1F. (B) Detalle del punto de contacto entre ambas proteínas. En rojo se representa una de las cadenas de pb1 y en beige la chaperona del fago K1F. El óvalo azul indica la posición en la que se encontraría la serina.

4.5.9 Unión al receptor

Previamente se ha descrito la interacción de la fibra del fago T5 con el antígeno O de los LPS, identificando como la unidad mínima para que se produzca la unión un trimanósido (Heller y Braun, 1982). Se ha llevado a cabo una búsqueda de los aminoácidos con los que más frecuentemente interactúa la manosa en las proteínas cuya estructura ha sido depositada con manosa como ligando, obteniendo como resultado que en la mayor parte de los casos esta interacción se lleva a cabo a través de asparagina, aspártico, serina o glicina. Además, en otras estructuras que interaccionan con la membrana externa, se ha visto que ésta tiene lugar frecuentemente a través de aminoácidos aromáticos. Si se analiza la posición de estos aminoácidos en la estructura, podemos destacar dos regiones del dominio aleta caudal, ambas situadas en los lazos que conectan las cadenas β , una de ellas Gly-Trp-Gly-Ser-Gly-Arg (aminoácidos 1245-1250, en el lazo Y-X), y la otra Asp-Asn-Tyr (aminoácidos 1239-1241, en el lazo W-Y) que aunque en el monómero no se encuentran cerca, en el trímero ambas regiones de dos monómeros diferentes sí se

encuentran muy próximas, generando por lo tanto una región que podría interaccionar con el receptor, como se representa en la Figura 4.30.

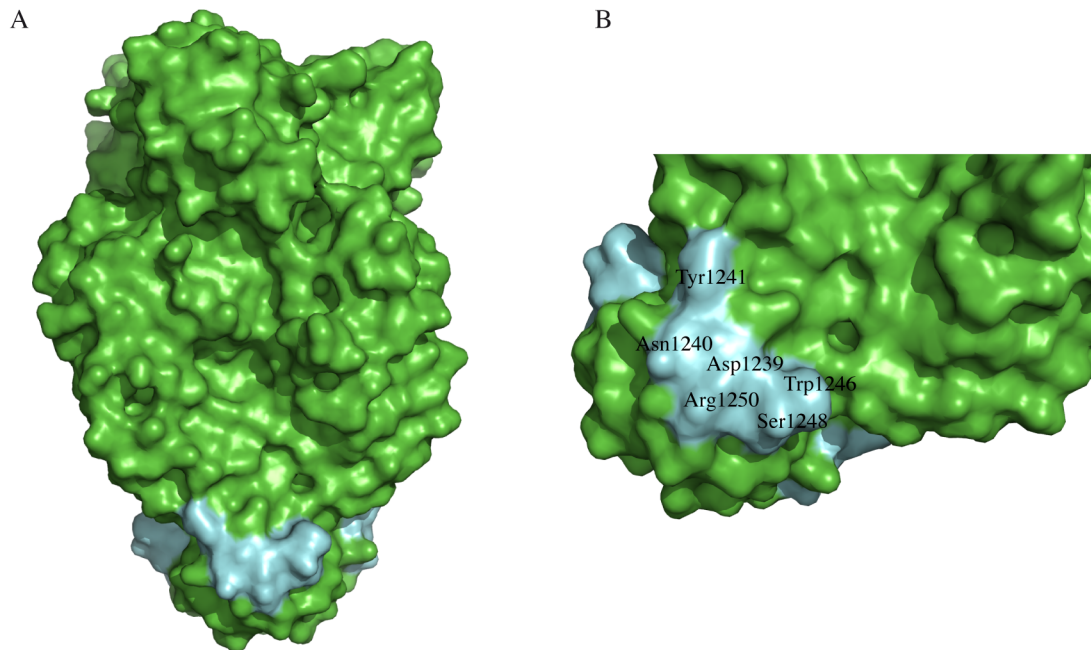


Figura 4.30: Aminoácidos identificados como posibles dianas de unión al receptor. (A) Representación de la superficie del fragmento resuelto de pb1, resaltando en cian los aminoácidos: 1239-1241 y 1245-1250. (B) Vista en detalle de la región en la que se localizan los aminoácidos resaltados. Se han etiquetado aquellos cuyas cadenas laterales podrían unirse al receptor.

4.5.10 Superposición de la estructura

Una vez que se obtuvo la estructura, se pudo realizar un ajuste de la misma en el interior del volumen obtenido por microscopía electrónica. Se obtuvo un ajuste en el cual la correlación final obtenida con el programa *Situs* es de 0,42 (Wriggers, 2010), que se presenta en la Figura 4.31. En la imagen se puede observar como la predicción del último dominio de la proteína, que era el que se había clonado en este fragmento, se ajusta bastante a lo obtenido por cristalografía, ya que la distribución de las cadenas es lineal y se cubre prácticamente la totalidad del volumen de este dominio.

Por otra parte, si se compara el volumen con el plegamiento de la proteína una vez ubicada en su interior, se puede observar que el extremo más distal se corresponde con el dominio aleta caudal de la proteína, que es más pequeño, y también que las

dos regiones en las que el dominio se estrecha (al principio y a la mitad) se corresponden con las dos regiones entrelazadas en la estructura de la proteína.

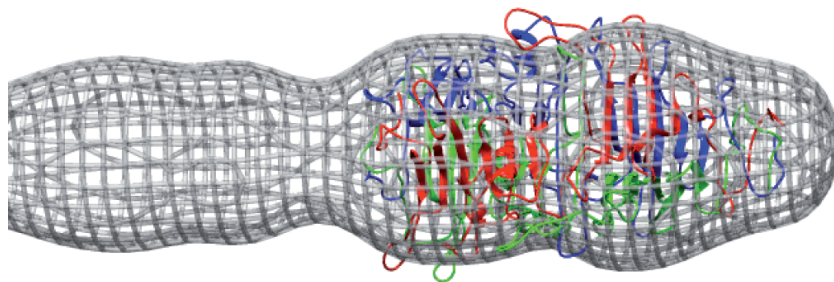


Figura 4.31: Ajuste de la estructura obtenida por cristalografía en el volumen obtenido por microscopía electrónica.

4.6 Comparación de *pb1* con otras estructuras

Para poder comparar la topología de este fragmento de la fibra de T5 con otras proteínas conocidas, se realizó una búsqueda con el servidor DALI. Las topologías más parecidas encontradas son las de las espinas de la cola de los fagos P22 y 9NA. Ambos fagos infectan *Salmonella*, pero pertenecen a familias diferentes ya que P22 es un podovirus y 9NA es un siphovirus. Los extremos carboxilo terminales de estas espinas se pueden superponer con la estructura de *pb1* con una diferencia (*rmsd*) de 4,4 Å en el caso de P22 (Figura 4.32) y 4.5 Å en el caso de 9NA. Como la estructura de 9NA fue descrita basándose en la de P22, se ha utilizado esta última para la comparación con la fibra de T5.

La estructura de la espina de P22 se divide en cuatro dominios: un dominio en β hélice paralela, una región entrecruzada, un dominio triangular en β -prisma y un dominio carboxilo terminal en forma de aleta caudal. Al igual que en el caso del fago 9NA, para *pb1* se ha utilizado una terminología similar para denominar a las distintas regiones, a excepción del primer dominio en β hélice paralela que no es similar en *pb1* ya que en *pb1* en esa región se localiza otro dominio triangular.

La espina de P22, al igual que la del fago 9NA tiene actividad endoramosidasa, degradando el antígeno O de los lipopolisacáridos de la pared celular. En el caso de T5, no se ha descrito que exista dicha actividad. Se han localizado los aminoácidos implicados en la actividad en P22 (359, 392 y 395) y se ha comparado su posición

con la estructura de pb1, observando que todos están localizados en lazos y que sólo el segundo se puede superponer en pb1, ya que los otros dos se encuentran en regiones que no están presentes en pb1, lo que concuerda con que esta actividad no se haya observado en T5. La otra región en la que se habían descrito residuos implicados en la actividad era en la aleta caudal, los residuos 489 y 505, pero esta región presenta un menor grado de homología. También se han localizado los residuos catalíticos del fago 9NA (375, 408 y 411) obteniendo los mismos resultados, por lo que se puede afirmar que pese a la similitud en la topología, al no coincidir la posición de los residuos catalíticos, es poco probable que T5 presentase actividad catalítica.

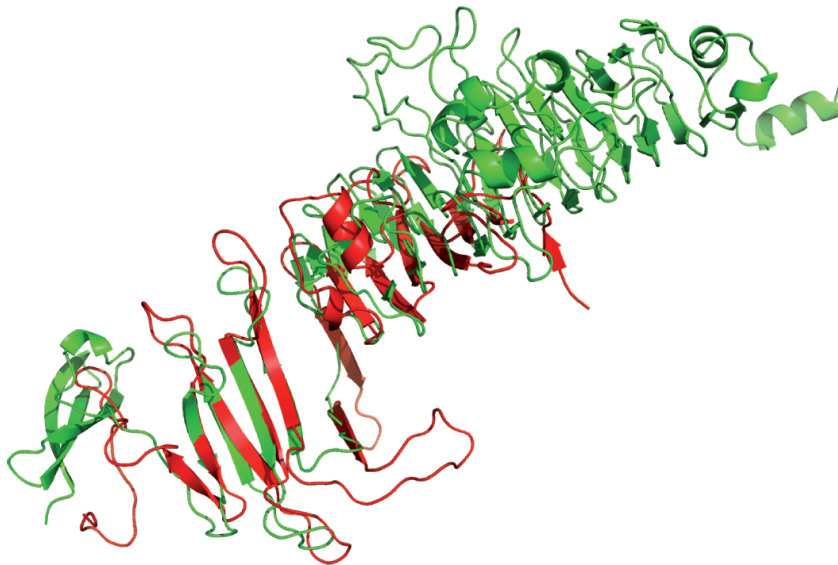


Figura 4.32: Superposición de la espina de la cola del fago P22 (verde) y la parte resuelta de pb1 (rojo). Se observa que las cadenas beta se pueden superponer en su mayor parte, encontrándose las diferencias principalmente en las conexiones entre dichas cadenas.

Comparando en más detalle la topología de ambas proteínas, se puede observar que la similitud se concentra en la región entrelazada y en el segundo dominio triangular. En la región entrelazada, la horquilla β en T5 está formada por dos cadenas β de seis aminoácidos cada una, pero en P22 las cadenas tienen siete y doce aminoácidos, y los lazos que las conectan también tienen tamaños muy dispares a los de la fibra de T5. En el caso del dominio triangular, en la espina de P22 las cadenas β tienden a ser de mayor tamaño que las equivalentes en la fibra de T5, pero los lazos

que las conectan son de menor longitud, siendo estos mucho más prominentes en la fibra de T5, dotando a la estructura de mayor flexibilidad. Es en estos lazos en los que se encuentran la mayor parte de los aminoácidos implicados en la actividad de la espina P22, y en el caso de T5 podrían estar también implicados en la interacción con el receptor.

Aunque la topología es claramente diferente, el tipo de plegamiento observado recuerda al que se obtuvo previamente en gp17 del fago T7, ya que se puede definir por la presencia de cadenas β , principalmente antiparalelas, que forman una lámina cóncava.

4.6.1 Alineamiento de secuencias

Tabla 4.10: Proteínas que presentan homología de secuencia con pb1 de T5.

Proteína	Nº de acceso	Identidad con pb1 de T5
Proteína hipotética del fago F20 (F20)	AEQ39205	74%
Fibra del profago de <i>E. coli</i> MS116 (MS116)	ZP_07162852	58%
Fibra del profago de <i>E. coli</i> MS175 (MS175)	ZP_07167889	58%
Fibra en <i>E. coli</i> H736 (H736)	ZP_08343077	58%
Fibra en <i>E. coli</i> HS (HS)	YP_001458341	58%
Proteína hipotética de <i>E. coli</i> KTE166 (KTE166)	ZP_203119116	57%
Fibra de profago de <i>E. coli</i> MS85 (MS85)	ZP_11539917	57%
Proteína hipotética EC970264 (EC970264)	ZP_23056309	57%
Proteína hipotética WCQ01161 (WCQ01161)	ZP_19619289	58%
Proteína hipotética WCQ01317 (WCQ01317)	ZP_19619443	58%

Se incluye el código de acceso de la base de datos de proteínas del NCBI y porcentaje de identidad. Entre paréntesis aparece la abreviatura que se utilizará posteriormente para el nombre de cada una de las proteínas.

Tras realizar la búsqueda de secuencias similares a pb1, se eligieron las que aparecen en la Tabla 4.10. Tras alinear las secuencias se observa que es en el extremo amino terminal donde se concentra la similitud entre todas las secuencias, a excepción de la proteína del fago F20. Las secuencias presentan homología con los

Estructura de fibras de bacteriófagos

primeros 600 aminoácidos de pb1 aproximadamente (Figura 4.33), región que según las predicciones realizadas en este trabajo abarcaría la región de unión a la cola del fago, toda la parte proximal de la fibra e incluso el comienzo de la parte distal.

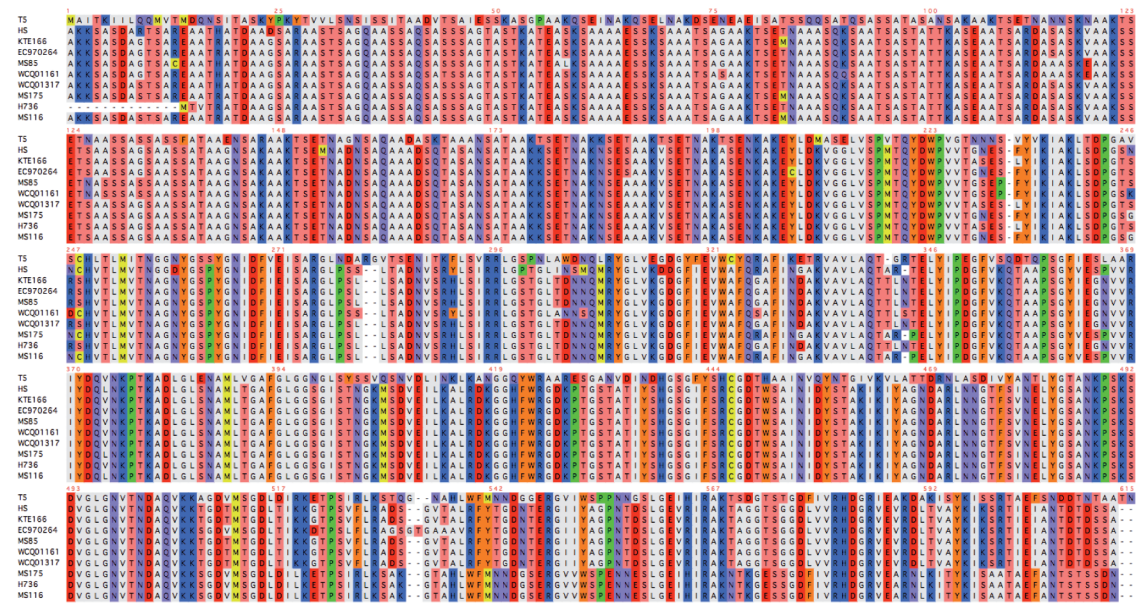


Figura 4.33: Alineamiento de secuencias de la fibra de T5 y sus homólogos a excepción del fago F20.

Por el contrario, cuando el alineamiento se hace exclusivamente entre la fibra de T5 con la proteína de F20 (Figura 4.34), la región en la que coinciden es en la carboxilo terminal, aproximadamente desde el aminoácido 950, aunque existen algunas coincidencias antes.

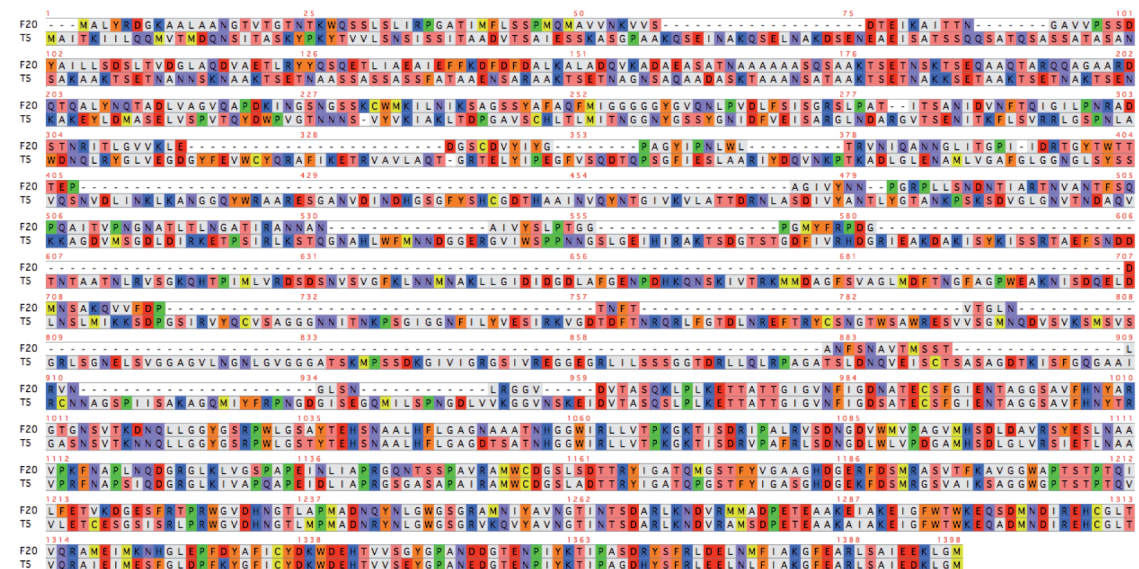


Figura 4.34: Alineamiento entre la secuencia de T5 y la de la proteína del fago F20.

Los resultados obtenidos al alinear la secuencia de la fibra de T5 con otras proteínas indicaron la existencia de cierto grado de conservación en la región amino terminal. Teniendo en cuenta la región de la fibra que abarca esta secuencia, se podría relacionar esta identidad de secuencia con la presente en otros muchos fagos que, por muy divergentes que sean sus fibras, siguen manteniendo homología en la zona de unión a la cola del fago. Hay que tener en cuenta que las secuencias que presentan homología en esta región son de proteínas poco estudiadas, siendo la mayor parte de proteínas localizadas en bacterias, principalmente en profagos, y que ninguna se ha analizado desde un punto de vista estructural.

Por otra parte, llama la atención la homología de secuencia localizada en el extremo carboxilo terminal entre la fibra de T5 y la proteína del fago F20, que es un siphovirus al igual que T5. Los aminoácidos que son homólogos son los correspondientes, según la predicción realizada en base a la microscopía, al último dominio de la fibra; más la parte correspondiente a la chaperona intramolecular que sería autoproteolizada. Esto indica claramente que la función de esta región en ambas proteínas debe ser la misma, y por lo tanto la proteína del fago F20 debe ser un apéndice de la cola implicado en la unión al receptor que presenta un dominio carboxilo terminal que se proteoliza, muy similar en secuencia al de T5, y probablemente en estructura a otros descritos anteriormente.

El fago F20 ha sido descrito recientemente como un siphovirus que infecta *Enterobacter aerogenes* (Mishra *et al.*, 2012). Aunque hasta el momento no se han realizado estudios estructurales con este fago, las diferencias existentes entre las dos secuencias podrían ser suficientes para modificar el rango de hospedador entre ambas especies bacterianas.

4.7 Perspectivas

El hecho de haber resuelto la estructura de la parte distal de gp17 ha abierto nuevas vías de investigación sobre la interacción del fago T7 con su célula hospedadora. Al conocer los aminoácidos que están expuestos en el extremo más distal y que podrían estar implicados en la unión al receptor, se han podido diseñar mutantes puntuales con los cuales se está analizando la implicación de estos aminoácidos en el proceso de unión a la célula hospedadora y también en el plegamiento de la proteína.

De la misma forma, la estructura de pb1 también abre nuevos caminos de investigación, tanto para la identificación de los aminoácidos que interaccionan con el receptor como para el posible intercambio de dominios con otras estructuras similares, dada la similitud con la espina de la cola del fago P22.

Tanto en gp17 como en pb1 se ha resuelto la estructura del extremo carboxilo terminal únicamente, pero en ambos casos se han encontrado plegamientos que no habían sido descritos previamente. En ambos casos se desconoce la estructura del resto de la proteína, que podría resultar interesante tanto desde el punto de vista estructural (nuevos plegamientos) como aplicado (la resistencia de las fibras es de gran interés en nanotecnología). Además, la estructura del extremo amino terminal ayudaría a aclarar cómo se produce la interacción entre las fibras y la cola del fago. Para ello, se podrían combinar las técnicas aquí descritas con otras como el dicroísmo circular o la expresión con proteínas de fusión. Por ejemplo, para obtener un fragmento intermedio de cualquiera de las dos fibras y que el fragmento mantuviese su naturaleza trimérica, sería necesario fusionarlo a un dominio de trimerización como el foldon de T4 (Letarov *et al.*, 1999).

Los avances en la comprensión de la interacción entre las fibras de los bacteriófagos y sus receptores son fundamentales para poder orientarlos a otros receptores, o incluso para poder utilizar a modo de sensores las propias fibras, y los resultados obtenidos en esta tesis pueden ayudar al desarrollo de aplicaciones tanto biotecnológicas como médicas, esperando que los bacteriófagos que habían quedado relegados a un segundo plano con el auge de los antibióticos, puedan ser una alternativa o complemento en la medicina moderna.

5 Conclusiones

1. Se ha clonado y expresado la proteína gp17 que forma la fibra del fago T7. Se ha determinado que la proteína se pliega correctamente sin necesidad de otras proteínas del fago y que la proteína obtenida tiene la misma forma que se había descrito anteriormente, pudiendo observarse sus rasgos característicos.
2. Se han clonado diferentes fragmentos carboxilo terminales y se ha comprobado el grado de expresión de todos ellos, eligiendo aquel que parecía más estable. Este fragmento, que abarca del aminoácido 371 al 553 de gp17, se purificó y se cristalizó en varias condiciones diferentes.
3. Con los datos obtenidos de una de las formas cristalinas se resolvió la estructura de la proteína gp17(371-553) a 1,9 Å, obteniendo las fases mediante Dispersión Anómala Múltiple. A continuación se pudo resolver otra de las formas cristalinas mediante reemplazamiento molecular a 2 Å.
4. Tras obtener la estructura de las dos formas cristalinas se pudieron comparar entre ellas, sin observarse diferencias significativas. Se definieron dos dominios principales en la estructura, un dominio pirámide y uno punta, que están conectados por una región que muestra flexibilidad entre las dos estructuras. Se comparó la estructura con las de otras proteínas, llegando a la conclusión de que el plegamiento de ambos dominios es nuevo, pero que existen otras proteínas fágicas con plegamientos similares.
5. Se purificaron colas del fago T5 y se aislaron imágenes de la parte distal de la proteína que forma las fibras, pb1, obteniendo un volumen por microscopía electrónica que permitió hacer una predicción de los diferentes dominios que forman la fibra.
6. Se clonaron varios fragmentos carboxilo terminales y se seleccionaron los dos más estables, pb1(794-1264) y pb1(970-1264) para proceder a la purificación y cristalización, obteniéndose cristales del fragmento pb1(970-1264) derivados con selenio-metionina, que permitieron resolver la estructura a 2,5 Å mediante Dispersión Anómala Simple.
7. Mediante la comparación con otras estructuras, se definieron los dominios principales de la estructura del fragmento pb1(993-1263), observando que existe cierta similitud con la espina del fago P22; y se analizó la posible interacción entre la proteína pb1 y la chaperona intramolecular carboxilo terminal empleando para ello como modelo la chaperona del fago K1F.

6 Bibliografía

- Abad-Zapatero C, Abdel-Meguid SS, Johnson JE, Leslie AG, Rayment I, Rossmann MG, Suck D y Tsukihara T (1980). Structure of southern bean mosaic virus at 2.8 Å resolution. *Nature* **286**(5768): 33-39.
- Abrahams JP y Leslie AGW (1996). Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Cryst D* **52**(1): 30-42.
- Abrescia NG, Bamford DH, Grimes JM y Stuart DI (2012). Structure unifies the viral universe. *Annu Rev Biochem* **81**: 795-822.
- Ackermann HW y Kropinski AM (2007). Curated list of prokaryote viruses with fully sequenced genomes. *Res Microbiol* **158**(7): 555-566.
- Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC y Zwart PH (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst D* **66**(2): 213-221.
- Agirrezabala X, Martin-Benito J, Caston JR, Miranda R, Valpuesta JM y Carrascosa JL (2005). Maturation of phage T7 involves structural modification of both shell and inner core components. *EMBO J* **24**(21): 3820-3829.
- Al-Zahrani AS, Kondabagil K, Gao S, Kelly N, Ghosh-Kumar M y Rao VB (2009). The small terminase, gp16, of bacteriophage T4 is a regulator of the DNA packaging motor. *J Biol Chem* **284**(36): 24490-24500.
- Allan G, Krakowka S, Ellis J y Charreyre C (2012). Discovery and evolving history of two genetically related but phenotypically different viruses, porcine circoviruses 1 and 2. *Virus Res* **164**(1-2): 4-9.
- Altschul SF, Gish W, Miller W, Myers EW y Lipman DJ (1990). Basic local alignment search tool. *J Mol Biol* **215**(3): 403-410.
- Audurier A y Martin C (1989). Phage typing of *Listeria monocytogenes*. *Int J Food Microbiol* **8**(3): 251-257.
- Bamford DH, Grimes JM y Stuart DI (2005). What does structure tell us about virus evolution? *Curr Opin Struct Biol* **15**(6): 655-663.
- Bartual SG, Garcia-Doval C, Alonso J, Schoehn G y van Raaij MJ (2010a). Two-chaperone assisted soluble expression and purification of the bacteriophage T4 long tail fibre protein gp37. *Protein Expr Purif* **70**(1): 116-121.

- Bartual SG, Otero JM, Garcia-Doval C, Llamas-Saiz AL, Kahn R, Fox GC y van Raaij MJ (2010b). Structure of the bacteriophage T4 long tail fiber receptor-binding tip. *Proc Natl Acad Sci U S A* **107**(47): 20287-20292.
- Battye TGG, Kontogiannis L, Johnson O, Powell HR y Leslie AGW (2011). iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Cryst D* **67**(4): 271-281.
- Beer M y Liu CQ (2012). Panning of a phage display library against a synthetic capsule for peptide ligands that bind to the native capsule of *Bacillus anthracis*. *PLoS One* **7**(9): e45472.
- Bergfors TM (2009). Protein Crystallization, International Univ Line.
- Bergh O, Borsheim KY, Bratbak G y Heldal M (1989). High abundance of viruses found in aquatic environments. *Nature* **340**(6233): 467-468.
- Bhardwaj A, Molineux IJ, Casjens SR y Cingolani G (2011). Atomic structure of bacteriophage sf6 tail needle knob. *J Biol Chem* **286**(35): 30867-30877.
- Bhattacharyya SP y Rao VB (1993). A novel terminase activity associated with the DNA packaging protein gp17 of bacteriophage T4. *Virology* **196**(1): 34-44.
- Blair JE y Williams RE (1961). Phage typing of staphylococci. *Bull World Health Organ* **24**(6): 771-784.
- Boman HG y Monner DA (1975). Characterization of lipopolysaccharides from *Escherichia coli* K-12 mutants. *J Bacteriol* **121**(2): 455-464.
- Boulanger P, Jacquot P, Plancon L, Chami M, Engel A, Parquet C, Herbeuval C y Letellier L (2008). Phage T5 straight tail fiber is a multifunctional protein acting as a tape measure and carrying fusogenic and muralytic activities. *J Biol Chem* **283**(20): 13556-13564.
- Boulanger P, le Maire M, Bonhivers M, Dubois S, Desmadril M y Letellier L (1996). Purification and structural and functional characterization of FhuA, a transporter of the *Escherichia coli* outer membrane. *Biochemistry* **35**(45): 14216-14224.
- Browning C, Shneider MM, Bowman VD, Schwarzer D y Leiman PG (2012). Phage pierces the host cell membrane with the iron-loaded spike. *Structure* **20**(2): 326-339.
- Brussow H, Canchaya C y Hardt WD (2004). Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol Mol Biol Rev* **68**(3): 560-602.

- Brussow H y Hendrix RW (2002). Phage genomics: small is beautiful. *Cell* **108**(1): 13-16.
- Buchan DW, Ward SM, Lobley AE, Nugent TC, Bryson K y Jones DT (2010). Protein annotation and modelling servers at University College London. *Nucleic Acids Res* **38**(Web Server issue): W563-568.
- Bujard H y Hendrickson HE (1973). Structure and function of the genome of coliphage T5. 1. The physical structure of the chromosome of T5 +. *Eur J Biochem* **33**(3): 517-528.
- Calendar RL y Abedon ST (2005). *The Bacteriophages*, Oxford University Press, USA.
- Campbell A (2003). The future of bacteriophage biology. *Nat Rev Genet* **4**(6): 471-477.
- Cerritelli ME, Cheng N, Rosenberg AH, McPherson CE, Booy FP y Steven AC (1997). Encapsidated conformation of bacteriophage T7 DNA. *Cell* **91**(2): 271-280.
- Cerritelli ME, Wall JS, Simon MN, Conway JF y Steven AC (1996). Stoichiometry and domainal organization of the long tail-fiber of bacteriophage T4: a hinged viral adhesin. *J Mol Biol* **260**(5): 767-780.
- Chang CY, Kemp P y Molineux IJ (2010). Gp15 and gp16 cooperate in translocating bacteriophage T7 DNA into the infected cell. *Virology* **398**(2): 176-186.
- Chen VB, Arendall WB, 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS y Richardson DC (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Cryst D* **66**(Pt 1): 12-21.
- Chung CT, Niemela SL y Miller RH (1989). One-step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution. *Proc Natl Acad Sci U S A* **86**(7): 2172-2175.
- Claverie JM, Abergel C y Ogata H (2009). Mimivirus. *Curr Top Microbiol Immunol* **328**: 89-121.
- Coombs DH y Arisaka F (1994). T4 Tail Structure and Function. Bacteriophage T4. J. D. Karam. Washington, DC, ASM Press.
- Cornelissen A, Ceyssens PJ, Krylov VN, Noben JP, Volckaert G y Lavigne R (2012). Identification of EPS-degrading activity within the tail spikes of the novel *Pseudomonas putida* phage AF. *Virology*.

- Cowtan K (2006). The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Cryst D* **62**(Pt 9): 1002-1011.
- d'Herelle F (1917). Sur Un Microbe Invisible Antagoniste Des Bacilles Dysentériques. *Comptes Rendus De L'Académie Des Sciences (París)* **165**(10): 373-375.
- Dawes J (1975). Characterisation of the bacteriophage T4 receptor site. *Nature* **256**(5513): 127-128.
- Demerec M y Fano U (1945). Bacteriophage-Resistant Mutants in *Escherichia coli*. *Genetics* **30**(2): 119-136.
- Dimasi N, Flot D, Dupeux F y Marquez JA (2007). Expression, crystallization and X-ray data collection from microcrystals of the extracellular domain of the human inhibitory receptor expressed on myeloid cells IREM-1. *Acta Cryst F* **63**(Pt 3): 204-208.
- Duckworth DH (1976). Who discovered bacteriophage? *Bacteriol Rev* **40**(4): 793-802.
- Effantin G, Boulanger P, Neumann E, Letellier L y Conway JF (2006). Bacteriophage T5 structure reveals similarities with HK97 and T4 suggesting evolutionary relationships. *J Mol Biol* **361**(5): 993-1002.
- Emsley P, Lohkamp B, Scott WG y Cowtan K (2010). Features and development of Coot. *Acta Cryst D* **66**(4): 486-501.
- Evans P (2006). Scaling and assessment of data quality. *Acta Cryst D* **62**(1): 72--82.
- Evans P (2011). An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Cryst D* **67**(4): 282-292.
- Fernandes S, Proenca D, Cantante C, Silva FA, Leandro C, Lourenco S, Milheirico C, de Lencastre H, Cavaco-Silva P, Pimentel M y Sao-Jose C (2012). Novel chimerical endolysins with broad antimicrobial activity against methicillin-resistant *Staphylococcus aureus*. *Microb Drug Resist* **18**(3): 333-343.
- Fischetti VA (2010). Bacteriophage endolysins: a novel anti-infective to control Gram-positive pathogens. *Int J Med Microbiol* **300**(6): 357-362.
- Fokine A, Zhang Z, Kanamaru S, Bowman VD, Aksyuk AA, Arisaka F, Rao VB y Rossmann MG (2013). The molecular architecture of the bacteriophage T4 neck. *J Mol Biol* **425**(10): 1731-1744.

- Fraser JS, Maxwell KL y Davidson AR (2007). Immunoglobulin-like domains on bacteriophage: weapons of modest damage? *Curr Opin Microbiol* **10**(4): 382-387.
- Frost JA, Kramer JM y Gillanders SA (1999). Phage typing of *Campylobacter jejuni* and *Campylobacter coli* and its use as an adjunct to serotyping. *Epidemiol Infect* **123**(1): 47-55.
- Garcia E, Elliott JM, Ramanculov E, Chain PS, Chu MC y Molineux IJ (2003). The genome sequence of *Yersinia pestis* bacteriophage phiA1122 reveals an intimate history with the coliphage T3 and T7 genomes. *J Bacteriol* **185**(17): 5248-5262.
- Garman E y Owen RL (2007). Cryocrystallography of macromolecules: practice and optimization. *Methods Mol Biol* **364**: 1-18.
- Grajewski BA, Kusek JW y Gelfand HM (1985). Development of a bacteriophage typing system for *Campylobacter jejuni* and *Campylobacter coli*. *J Clin Microbiol* **22**(1): 13-18.
- Guardado Calvo P, Fox GC, Hermo Parrado XL, Llamas-Saiz AL, Costas C, Martinez-Costas J, Benavente J y van Raaij MJ (2005). Structure of the carboxy-terminal receptor-binding domain of avian reovirus fibre sigmaC. *J Mol Biol* **354**(1): 137-149.
- Hankin EH (1896). L'action bactericide des eaux de la Jumna et du Gange sur le vibron du cholera. *Annales de l'Institut Pasteur* **10**: 511-523.
- Harrison SC, Olson AJ, Schutt CE, Winkler FK y Bricogne G (1978). Tomato bushy stunt virus at 2.9 Å resolution. *Nature* **276**(5686): 368-373.
- Heineman RH, Springman R y Bull JJ (2008). Optimal foraging by bacteriophages through host avoidance. *Am Nat* **171**(4): E149-157.
- Heller K y Braun V (1982). Polymannose O-antigens of *Escherichia coli*, the binding sites for the reversible adsorption of bacteriophage T5+ via the L-shaped tail fibers. *J Virol* **41**(1): 222-227.
- Heller KJ (1992). Molecular interaction between bacteriophage and the gram-negative cell envelope. *Arch Microbiol* **158**(4): 235-248.
- Heller KJ y Bryniok D (1984). O antigen-dependent mutant of bacteriophage T5. *J Virol* **49**(1): 20-25.

- Heller KJ y Schwarz H (1985). Irreversible binding to the receptor of bacteriophages T5 and BF23 does not occur with the tip of the tail. *J Bacteriol* **162**(2): 621-625.
- Hendrickson WA, Horton JR y LeMaster DM (1990). Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J* **9**(5): 1665-1672.
- Heymann JB (2001). Bsoft: image and molecular processing in electron microscopy. *J Struct Biol* **133**(2-3): 156-169.
- Heymann JB y Belnap DM (2007). Bsoft: image processing and molecular modeling for electron microscopy. *J Struct Biol* **157**(1): 3-18.
- Holm L y Rosenstrom P (2010). Dali server: conservation mapping in 3D. *Nucleic Acids Res* **38**(Web Server issue): W545-549.
- Hu B, Margolin W, Molineux IJ y Liu J (2013). The bacteriophage τ 7 virion undergoes extensive structural remodeling during infection. *Science* **339**(6119): 576-579.
- Ishii T, Yamaguchi Y y Yanagida M (1978). Binding of the structural protein soc to the head shell of bacteriophage T4. *J Mol Biol* **120**(4): 533-544.
- Jain P, Hartman TE, Eisenberg N, O'Donnell MR, Kriakov J, Govender K, Makume M, Thaler DS, Hatfull GF, Sturm AW, Larsen MH, Moodley P y Jacobs WR, Jr. (2012). phi(2)GFP10, a high-intensity fluorophage, enables detection and rapid drug susceptibility testing of *Mycobacterium tuberculosis* directly from sputum samples. *J Clin Microbiol* **50**(4): 1362-1369.
- Jones DT (1999). Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* **292**(2): 195-202.
- Kabsch W (2010). XDS. *Acta Cryst D* **66**(2): 125--132.
- Kanamaru S, Leiman PG, Kostyuchenko VA, Chipman PR, Mesyanzhinov VV, Arisaka F y Rossmann MG (2002). Structure of the cell-puncturing device of bacteriophage T4. *Nature* **415**(6871): 553-557.
- Kemp P, Garcia LR y Molineux IJ (2005). Changes in bacteriophage T7 virion structure at the initiation of infection. *Virology* **340**(2): 307-317.
- Kleywegt GJ (2000). Validation of protein crystal structures. *Acta Cryst D* **56**(Pt 3): 249-265.

- Kostyuchenko VA, Leiman PG, Chipman PR, Kanamaru S, van Raaij MJ, Arisaka F, Mesyanzhinov VV y Rossmann MG (2003). Three-dimensional structure of bacteriophage T4 baseplate. *Nat Struct Biol* **10**(9): 688-693.
- Kreisberg JF, Betts SD, Haase-Pettingell C y King J (2002). The interdigitated beta-helix domain of the P22 tailspike protein acts as a molecular clamp in trimer stabilization. *Protein Sci* **11**(4): 820-830.
- Krissinel E y Henrick K (2007). Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**(3): 774-797.
- Krupovic M, Prangishvili D, Hendrix RW y Bamford DH (2011). Genomics of bacterial and archaeal viruses: dynamics within the prokaryotic virosphere. *Microbiol Mol Biol Rev* **75**(4): 610-635.
- Laemmli UK (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**(5259): 680-685.
- Langer G, Cohen SX, Lamzin VS y Perrakis A (2008). Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat Protoc* **3**(7): 1171-1179.
- Lanni YT (1958). Lysis inhibition with a mutant of bacteriophage T5. *Virology* **5**(3): 481-501.
- Lanni YT (1968). First-step-transfer deoxyribonucleic acid of bacteriophage T5. *Bacteriol Rev* **32**(3): 227-242.
- Leiman PG, Arisaka F, van Raaij MJ, Kostyuchenko VA, Aksyuk AA, Kanamaru S y Rossmann MG (2010). Morphogenesis of the T4 tail and tail fibers. *Viol J* **7**: 355.
- Leiman PG, Kanamaru S, Mesyanzhinov VV, Arisaka F y Rossmann MG (2003). Structure and morphogenesis of bacteriophage T4. *Cell Mol Life Sci* **60**(11): 2356-2370.
- Leiman PG y Shneider MM (2012). Contractile tail machines of bacteriophages. *Adv Exp Med Biol* **726**: 93-114.
- Letarov AV, Londer YY, Boudko SP y Mesyanzhinov VV (1999). The carboxy-terminal domain initiates trimerization of bacteriophage T4 fibrin. *Biochemistry (Mosc)* **64**(7): 817-823.
- Marabini R, Masegosa IM, San Martin MC, Marco S, Fernandez JJ, de la Fraga LG, Vaquerizo C y Carazo JM (1996). Xmipp: An Image Processing Package for Electron Microscopy. *J Struct Biol* **116**(1): 237-240.

- Marzari R, Sblattero D, Righi M y Bradbury A (1997). Extending filamentous phage host range by the grafting of a heterologous receptor binding domain. *Gene* **185**(1): 27-33.
- McCorquodale DJ, Shaw AR, Shaw PK y Chinnadurai G (1977). Pre-early polypeptides of bacteriophages T5 and BF23. *J Virol* **22**(2): 480-488.
- McPherson A (2004). Introduction to protein crystallization. *Methods* **34**(3): 254-265.
- Meng X, Shi Y, Ji W, Meng X, Zhang J, Wang H, Lu C, Sun J y Yan Y (2011). Application of a bacteriophage lysin to disrupt biofilms formed by the animal pathogen *Streptococcus suis*. *Appl Environ Microbiol* **77**(23): 8272-8279.
- Merckel MC, Huiskonen JT, Bamford DH, Goldman A y Tuma R (2005). The structure of the bacteriophage PRD1 spike sheds light on the evolution of viral capsid architecture. *Mol Cell* **18**(2): 161-170.
- Mesyanzhinov VV (2004). Bacteriophage T4: structure, assembly, and initiation infection studied in three dimensions. *Adv Virus Res* **63**: 287-352.
- Miernikiewicz P, Owczarek B, Piotrowicz A, Boczkowska B, Rzewucka K, Figura G, Letarov A, Kulikov E, Kopciuch A, Switala-Jelen K, Oslizlo A, Hodyra K, Gubernator J y Dabrowska K (2012). Recombinant Expression and Purification of T4 Phage Hoc, Soc, gp23, gp24 Proteins in Native Conformations with Stability Studies. *PLoS One* **7**(7): e38902.
- Mindell JA y Grigorieff N (2003). Accurate determination of local defocus and specimen tilt in electron microscopy. *J Struct Biol* **142**(3): 334-347.
- Mishra CK, Choi TJ y Kang SC (2012). Isolation and characterization of a bacteriophage F20 virulent to *Enterobacter aerogenes*. *J Gen Virol* **93**(Pt 10): 2310-2314.
- Mitraki A, Barge A, Chroboczek J, Andrieu JP, Gagnon J y Ruigrok RW (1999). Unfolding studies of human adenovirus type 2 fibre trimers. Evidence for a stable domain. *Eur J Biochem* **264**(2): 599-606.
- Moffat K y Ren Z (1997). Synchrotron radiation applications to macromolecular crystallography. *Curr Opin Struct Biol* **7**(5): 689-696.
- Molineux IJ (2001). No syringes please, ejection of phage T7 DNA from the virion is enzyme driven. *Mol Microbiol* **40**(1): 1-8.

- Monjezi R, Tan SW, Tey BT, Sieo CC y Tan WS (2013). Detection of hepatitis B virus core antigen by phage display mediated TaqMan real-time immuno-PCR. *J Virol Methods* **187**(1): 121-126.
- Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F y Vagin AA (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Cryst D* **67**(4): 355-367.
- Novick RP, Christie GE y Penades JR (2010). The phage-related chromosomal islands of Gram-positive bacteria. *Nat Rev Microbiol* **8**(8): 541-551.
- Novy R, Yaeger K, Held D y Mierendorf R (2002). Coexpression of multiple target proteins in *E. coli*. *Innovations* **12**: 2-6.
- Ohnishi M, Kurokawa K y Hayashi T (2001). Diversification of *Escherichia coli* genomes: are bacteriophages the major contributors? *Trends Microbiol* **9**(10): 481-485.
- Olia AS, Casjens S y Cingolani G (2007). Structure of phage P22 cell envelope-penetrating needle. *Nat Struct Mol Biol* **14**(12): 1221-1226.
- Orlova E (2012). Bacteriophages and Their Structural Organisation. Bacteriophages. I. Kurtboke, InTech.
- Pande J, Szewczyk MM y Grover AK (2010). Phage display: concept, innovations, applications and future. *Biotechnol Adv* **28**(6): 849-858.
- Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC y Ferrin TE (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**(13): 1605-1612.
- Pierce BG, Hourai Y y Weng Z (2011). Accelerating protein docking in ZDOCK using an advanced 3D convolution library. *PLoS One* **6**(9): e24657.
- Qimron U, Marintcheva B, Tabor S y Richardson CC (2006). Genomewide screens for *Escherichia coli* genes affecting growth of T7 bacteriophage. *Proc Natl Acad Sci U S A* **103**(50): 19039-19044.
- Raetz CR (1990). Biochemistry of endotoxins. *Annu Rev Biochem* **59**: 129-170.
- Raetz CR y Whitfield C (2002). Lipopolysaccharide endotoxins. *Annu Rev Biochem* **71**: 635-700.
- Ramachandran GN, Ramakrishnan C y Sasisekharan V (1963). Stereochemistry of polypeptide chain configurations. *J Mol Biol* **7**: 95-99.
- Rao VB y Black LW (2010). Structure and assembly of bacteriophage T4 head. *Virology* **7**: 356.

- Rossmann MG (1990). The molecular replacement method. *Acta Cryst A* **46** (Pt 2): 73-82.
- Rossmann MG, Arnold E, Erickson JW, Frankenberger EA, Griffith JP, Hecht HJ, Johnson JE, Kamer G, Luo M, Mosser AG y et al. (1985). Structure of a human common cold virus and functional relationship to other picornaviruses. *Nature* **317**(6033): 145-153.
- Rossmann MG y Blow DM (1962). The detection of sub-units within the crystallographic asymmetric unit. *Acta Cryst* **15**(1): 24-31.
- Ruiz N, Kahne D y Silhavy TJ (2009). Transport of lipopolysaccharide across the cell envelope: the long road of discovery. *Nat Rev Microbiol* **7**(9): 677-683.
- Rupp B (2009). *Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology*. New York, Garland Science.
- Sathaliyawala T, Islam MZ, Li Q, Fokine A, Rossmann MG y Rao VB (2010). Functional analysis of the highly antigenic outer capsid protein, Hoc, a virus decoration protein from T4-like bacteriophages. *Mol Microbiol* **77**(2): 444-455.
- Schägger H (2006). Tricine-SDS-PAGE. *Nat Protoc* **1**(1): 16-22.
- Scheres SH, Valle M y Carazo JM (2005). Fast maximum-likelihood refinement of electron microscopy images. *Bioinformatics* **21 Suppl 2**: ii243-244.
- Schofield DA, Molineux IJ y Westwater C (2012). Rapid identification and antibiotic susceptibility testing of *Yersinia pestis* using bioluminescent reporter phage. *J Microbiol Methods* **90**(2): 80-82.
- Schulz EC, Dickmanns A, Urlaub H, Schmitt A, Muhlenhoff M, Stummeyer K, Schwarzer D, Gerardy-Schahn R y Ficner R (2010a). Crystal structure of an intramolecular chaperone mediating triple-beta-helix folding. *Nat Struct Mol Biol* **17**(2): 210-215.
- Schulz EC, Schwarzer D, Frank M, Stummeyer K, Muhlenhoff M, Dickmanns A, Gerardy-Schahn R y Ficner R (2010b). Structural basis for the recognition and cleavage of polysialic acid by the bacteriophage K1F tailspike protein EndoNF. *J Mol Biol* **397**(1): 341-351.
- Schwarzer D, Stummeyer K, Gerardy-Schahn R y Muhlenhoff M (2007). Characterization of a novel intramolecular chaperone domain conserved in endosialidases and other bacteriophage tail spike and fiber proteins. *J Biol Chem* **282**(5): 2821-2831.

- Serwer P (1976). Internal proteins of bacteriophage T7. *J Mol Biol* **107**(3): 271-291.
- Serwer P, Khan SA, Hayes SJ, Watson RH y Griess GA (1997). The conformation of packaged bacteriophage T7 DNA: informative images of negatively stained T7. *J Struct Biol* **120**(1): 32-43.
- Serwer P, Wright ET, Hakala KW y Weintraub ST (2008). Evidence for bacteriophage T7 tail extension during DNA injection. *BMC Res Notes* **1**: 36.
- Seydel U, Schroder G y Brandenburg K (1989). Reconstitution of the lipid matrix of the outer membrane of gram-negative bacteria as asymmetric planar bilayer. *J Membr Biol* **109**(2): 95-103.
- Sheldrick G (2008). A short history of SHELX. *Acta Cryst A* **64**(1): 112-122.
- Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD y Higgins DG (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**: 539.
- Snyder CE, Jr. y Benzinger RH (1981). Second-step transfer of bacteriophage T5 DNA: purification and characterization of the T5 gene A2 protein. *J Virol* **40**(1): 248-257.
- Spinelli S, Campanacci V, Blangy S, Moineau S, Tegoni M y Cambillau C (2006a). Modular structure of the receptor binding proteins of *Lactococcus lactis* phages. The RBP structure of the temperate phage TP901-1. *J Biol Chem* **281**(20): 14256-14262.
- Spinelli S, Desmyter A, Verrips CT, de Haard HJ, Moineau S y Cambillau C (2006b). Lactococcal bacteriophage p2 receptor-binding protein structure suggests a common ancestor gene with bacterial and mammalian viruses. *Nat Struct Mol Biol* **13**(1): 85-89.
- Stanley WM (1935). Isolation of a Crystalline Protein Possessing the Properties of Tobacco-Mosaic Virus. *Science* **81**(2113): 644-645.
- Steven AC, Heymann JB, Cheng N, Trus BL y Conway JF (2005). Virus maturation: dynamics and mechanism of a stabilizing structural transition that leads to infectivity. *Curr Opin Struct Biol* **15**(2): 227-236.
- Steven AC, Trus BL, Maizel JV, Unser M, Parry DA, Wall JS, Hainfeld JF y Studier FW (1988). Molecular substructure of a viral receptor-recognition protein. The gp17 tail-fiber of bacteriophage T7. *J Mol Biol* **200**(2): 351-365.

- Studier FW, Rosenberg AH, Dunn JJ y Dubendorff JW (1990). Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol* **185**: 60-89.
- Stummeyer K, Dickmanns A, Muhlenhoff M, Gerardy-Schahn R y Ficner R (2005). Crystal structure of the polysialic acid-degrading endosialidase of bacteriophage K1F. *Nat Struct Mol Biol* **12**(1): 90-96.
- Summers WC (2001). Bacteriophage therapy. *Annu Rev Microbiol* **55**: 437-451.
- Tang G, Peng L, Baldwin PR, Mann DS, Jiang W, Rees I y Ludtke SJ (2007). EMAN2: an extensible image processing suite for electron microscopy. *J Struct Biol* **157**(1): 38-46.
- Thomassen E, Gielen G, Schutz M, Schoehn G, Abrahams JP, Miller S y van Raaij MJ (2003). The structure of the receptor-binding domain of the bacteriophage T4 short tail fibre reveals a knitted trimeric metal-binding fold. *J Mol Biol* **331**(2): 361-373.
- Tripathi P, Beaussart A, Andre G, Rolain T, Lebeer S, Vanderleyden J, Hols P y Dufrene YF (2012). Towards a nanoscale view of lactic acid bacteria. *Micron* **43**(12): 1323-1330.
- Twort FW (1915). An Investigation on the Nature of Ultramicroscopic Viruses. *Lancet* **189**(2): 1241-1243.
- Vagin AA y Isupov MN (2001). Spherically averaged phased translation function and its application to the search for molecules and fragments in electron-density maps. *Acta Cryst D* **57**(10): 1451-1456.
- van Raaij MJ, Mitraki A, Lavigne G y Cusack S (1999). A triple beta-spiral in the adenovirus fibre shaft reveals a new structural motif for a fibrous protein. *Nature* **401**(6756): 935-938.
- Van Regenmortel MHV, Fauquet CM, Bishop DHL, Cartens E, Estes MK, Lemon S, NManiloff J, Mayo MA, McGeoch DJ, Pringle CR y Wickner R (2000). *Virus Taxonomy: Classification and Nomenclature of Viruses : Seventh Report of the International Committee on Taxonomy of Viruses*. San Diego, Academic Press.
- Veesler D y Cambillau C (2011). A common evolutionary origin for tailed-bacteriophage functional modules and bacterial machineries. *Microbiol Mol Biol Rev* **75**(3): 423-433.

- Vonrhein C, Blanc E, Roversi P y Bricogne G (2007). Automated structure solution with autoSHARP. *Methods Mol Biol* **364**: 215-230.
- Walsh MA, Evans G, Sanishvili R, Dementieva I y Joachimiak A (1999). MAD data collection - current trends. *Acta Cryst D* **55**(Pt 10): 1726-1732.
- Waseh S, Hanifi-Moghaddam P, Coleman R, Masotti M, Ryan S, Foss M, MacKenzie R, Henry M, Szymanski CM y Tanha J (2010). Orally administered P22 phage tailspike protein reduces *Salmonella* colonization in chickens: prospects of a novel therapy against bacterial infections. *PLoS One* **5**(11): e13904.
- Westwater C, Kasman LM, Schofield DA, Werner PA, Dolan JW, Schmidt MG y Norris JS (2003). Use of genetically engineered phage to deliver antimicrobial agents to bacteria: an alternative therapy for treatment of bacterial infections. *Antimicrob Agents Chemother* **47**(4): 1301-1307.
- Winter G (2010). *xia2*: an expert system for macromolecular crystallography data reduction. *J Appl Cryst* **43**(1).
- Wood WB y Henninger M (1969). Attachment of tail fibers in bacteriophage T4 assembly: some properties of the reaction *in vitro* and its genetic control. *J Mol Biol* **39**(3): 603-618.
- Wriggers W (2010). Using Situs for the integration of multi-resolution structures. *Biophys Rev* **2**(1): 21-27.
- Xiang Y, Leiman PG, Li L, Grimes S, Anderson DL y Rossmann MG (2009). Crystallographic insights into the autocatalytic assembly mechanism of a bacteriophage tail spike. *Mol Cell* **34**(3): 375-386.
- Yamashita E, Nakagawa A, Takahashi J, Tsunoda K, Yamada S y Takeda S (2011). The host-binding domain of the P2 phage tail spike reveals a trimeric iron-binding structure. *Acta Cryst F* **67**(Pt 8): 837-841.
- Yu F y Mizushima S (1982). Roles of lipopolysaccharide and outer membrane protein OmpC of *Escherichia coli* K-12 in the receptor function for bacteriophage T4. *J Bacteriol* **151**(2): 718-722.
- Zweig M y Cummings DJ (1973). Structural proteins of bacteriophage T5. *Virology* **51**(2): 443-453.

Anexo 1: Tabla de oligonucleótidos

Nombre	Secuencia
gp17.F.NcoI	CCGGCCATGGCTAACTAATTAACCGTT
gp17.R.HindIII	CCGGAAGCTTACTCGTTCTCCACCATGATTG
gp17.F281.NcoI	CCGGCCATGGGAACCAATGTGTACTGAAAG
gp17.F371.NcoI	CCGGCCATGGGGCATGTTCTTCAATTAGAGTC
gp17.F431.NcoI	CCGGCCATGGGACACTTCCACGTAGGTCAGG
gp17.F461.NcoI	CCGGCCATGGACAGCTTCGTTGCGAAAGTC
pb1.F1.BamHI	GGTCGGATCCTGGCTATAACTAAAATAATTCTACAGC
pb1.F224.BamHI	GGTCGGATCCCCGTCGGTACAAACAACAATAGCG
pb1.F413.BamHI	GGTCGGATCCTTAAAGCTAATGGCGGTCAATACTGG
pb1.F674.BamHI	GGTCGGATCCTGGATGCTGGTTTCTCTGTTGCTGG
pb1.F693.BamHI	GGTCGGATCCCTTGGGAGGCTAAAAAATTAGTGACC
pb1.F794.BamHI	GGTCGGATCCTGAACCAAGATGTGAGTGTTAAGTCAATG
pb1.F1106.BamHI	GGTCGGATCCTAAATGCTGCTGTGCCTAGGTTTAACG
pb1.F728.BamHI	GCGCGGATCCGGTGGCGGCAACAACATTAC
pb1.F835.BamHI	GCGCGGATCCAAAATGCCATCGTCTGATAAGG
pb1.F970.BamHI	GCGCGGATCCGAGACAACCGCGACAACC
pb1.R.NotI	ATAAGAATGCGGCCGCTTACATACCTAATTTATCCTCAATGC
pb1.R.HindIII	GATCCCAAGCTTTTACATACCTAATTTATCCTCAATTGC
T7 promoter primer	TAATACGACTCACTATAGGG
T7 terminator primer	GCTAGTTATTGCTCAGCGG

Anexo 2: Índice de figuras

Figura 1.1: Virus pertenecientes a los cuatro linajes descritos según la similitud de sus proteínas de la cápside.....	14
Figura 1.2: Imágenes de microscopía electrónica de representantes de cada una de las tres familias de Caudovirales.	16
Figura 1.3: Ciclo de replicación de un fago..	18
Figura 1.4: Esquema de la pared celular de bacterias Gram positivas y Gram negativas.....	19
Figura 1.5: Esquema de la estructura de los lipopolisacáridos.....	20
Figura 1.6: El myovirus T4.	23
Figura 1.7: El siphovirus T5.....	25
Figura 1.8: El podovirus T7.	28
Figura 3.1: Principales tipos de técnicas cromatográficas.	46
Figura 3.2: Técnicas empleadas para llevar a cabo el proceso de cristalización.....	48
Figura 3.3: La recogida de datos de difracción.	49
Figura 3.4: El problema de la fase.....	51
Figura 3.5: Esquema de la fibra de T7.	54
Figura 4.1: Modelos de vectores construidos para la expresión de gp17 o sus fragmentos.	71
Figura 4.2: Purificación de gp17 con cola de histidinas.....	74
Figura 4.3: Purificación de gp17(371-553) con cola de histidinas.....	75
Figura 4.4: Tripsinización de gp17 y gp17(371-553).....	76
Figura 4.5: Imagen de fibras de T7 obtenida mediante microscopía electrónica.....	77
Figura 4.6: Cristales de gp17(371-553).....	78
Figura 4.7: Patrones de difracción de cristales de gp17(371-553).	79
Figura 4.8: Estructura de gp17(371-553).	83
Figura 4.9: Representación de la superficie de la gp17(371-553) en función de la carga de sus aminoácidos.	84
Figura 4.10: Conservación de secuencia en gp17(371-553).	88
Figura 4.11: Inhibición de la formación de placas de lisis mediante la unión a LPS.....	91
Figura 4.12: Alineamiento de las secuencias completas de las fibras de YpsP-R, phiA1122, YpsP-Y, T3, YpsP-G, 13a y T7.....	95

Figura 4.13: Representación en escala de grises de la conservación de los aminoácidos en la secuencia del fragmento gp17(371-553).	97
Figura 4.14: Comparación de la estructura del dominio punta de gp17(371-553) con otras estructuras.	98
Figura 4.15: Comparación de la estructura del dominio pirámide de gp17(371-553) con otras estructuras.	99
Figura 4.16: Proceso de purificación de las colas del fago T5.....	100
Figura 4.17: Microscopía electrónica de las fibras de la cola de T5.	101
Figura 4.18: Predicción de estructura secundaria de pb1.	104
Figura 4.19: Modelo de vector obtenido para uno de los fragmentos de pb1.	105
Figura 4.20: Muestras de expresión y/o purificación de cada uno de los fragmentos de pb1 clonados.	106
Figura 4.21: Purificación del fragmento pb1(794-1264).....	107
Figura 4.22: Purificación del fragmento pb1(1106-1264).....	108
Figura 4.23: Purificación del fragmento pb1(970-1264).....	109
Figura 4.24: Diferentes morfologías obtenidas en los cristales del fragmento pb1(794-1264)..	110
Figura 4.25: Cristales obtenidos del fragmento pb1(970-1264).....	111
Figura 4.26: Patrones de difracción obtenidos para los dos fragmentos de pb1..	112
Figura 4.27: Estructura de pb1(993-1263).	114
Figura 4.28: Representación de las diferentes cargas presentes en pb1.	116
Figura 4.29: Interacción entre la proteína pb1 y la chaperona carboxilo terminal...119	
Figura 4.30: Aminoácidos identificados como posibles dianas de unión al receptor.	120
Figura 4.31: Ajuste de la estructura obtenida por cristalografía en el volumen obtenido por microscopía electrónica.....	121
Figura 4.32: Superposición de la espina de la cola del fago P22 y la parte resuelta de pb1	122
Figura 4.33: Alineamiento de secuencias de la fibra de T5 y sus homólogos a excepción del fago F20.....	124
Figura 4.34: Alineamiento entre la secuencia de T5 y la del fago F20.....	124

Anexo 3: Índice de tablas

Tabla 3.1: Oligonucleótidos necesarios para amplificar cada uno de los fragmentos diseñados de pb1.	64
Tabla 4.1: : Estadísticas obtenidas del escalado de los sets de datos nativos de gp17(371-553).	79
Tabla 4.2: : Estadísticas obtenidas del escalado de los sets de datos utilizados para el MAD de gp17(371-553).	80
Tabla 4.3: Estadísticas del refinamiento de gp17(371-553).	80
Tabla 4.4: Estadísticas del faseado de gp17(371-553).	81
Tabla 4.5: Ensayos de inhibición de la infección del fago T7.	92
Tabla 4.6: Fibras que presentan homología de secuencia con gp17 de T7.	94
Tabla 4.7: Estadísticas obtenidas del escalado del sets de datos utilizado para el SAD.	111
Tabla 4.8: Estadísticas del faseado	112
Tabla 4.9: Estadísticas del refinamiento.	113
Tabla 4.10: Proteínas que presentan homología de secuencia con pb1 de T5.	123

Anexo 4: Artículos publicados durante el desarrollo de la tesis

- Artículos publicados cuyos resultados forman parte de esta tesis:

Garcia-Doval C y van Raaij MJ (2012). Crystallization of the C-terminal domain of the bacteriophage T7 fibre protein gp17. *Acta Crystallographica Section F* **68**(2): 166-171.

Garcia-Doval C y van Raaij MJ (2012). Structure of the receptor-binding carboxy-terminal domain of bacteriophage T7 tail fibers. *Proc Natl Acad Sci U S A* **109**(24): 9390-9395.

- Otros artículos:

Bartual SG, Garcia-Doval C, Alonso J, Schoehn G y van Raaij MJ (2010). Two-chaperone assisted soluble expression and purification of the bacteriophage T4 long tail fibre protein gp37. *Protein Expr Purif* **70**(1): 116-121.

Bartual SG, Otero JM, Garcia-Doval C, Llamas-Saiz AL, Kahn R, Fox GC y van Raaij MJ (2010). Structure of the bacteriophage T4 long tail fiber receptor-binding tip. *Proc Natl Acad Sci U S A* **107**(47): 20287-20292.

